NISS

Formulation and Solution of a Dynamic User-Optimal Route Choice Model on a Large-Scale Traffic Network

Der-Horng Lee, David E. Boyce, and Bruce N. Janson

Summary

This research presents a dynamic user-optimal (DUO) route choice model for predicting dynamic traffic conditions, intended for off-line Advanced Traffic Management Systems (ATMS) and Advanced Traveler Information Systems (ATIS) evaluation and implementation. This DUO route choice model is formulated as a variational inequality (VI) and can be solved efficiently to convergence by the proposed diagonalization algorithm with discrete time intervals.

The test network selected for testing the proposed dynamic user-optimal (DUO) route choice model is the ADVANCE Network. The ADVANCE Network is located in the northwestern suburbs of Chicago and covers about 300 square miles (800 square kilometers). To generate specific link travel times for the investigated network, the expanded intersection representation is employed. Using this network representation, each turning movement is coded as an individual intersection link, increasing the network scale to approximately three times larger than the conventional network representation. Nearly 10,000 nodes and 23,000 links are defined for the solution procedure. For most links a realistic traffic engineering-based link travel time function, the Akcelik function, is adopted in this research, in place of the simplistic but widely used BPR (Bureau of Public Roads) function, to estimate delays and travel times for various types of links and intersections. Unexpected capacity reducing events causing nonrecurrent traffic congestion are analyzed with the model. Route choice behavior based on anticipatory and non-anticipatory network conditions are considered in performing the incident analysis, extending the capability of this model to contribute to the evaluation of ATMS and ATIS.

Four global network performance measures and a convergence index are defined to monitor the solution process of the model and assess the dynamic traffic condition over the ADVANCE Network. Results from different locations within the ADVANCE Network under different incident scenarios are analyzed in detail. Although not yet fully validated, this model is able to predict time-dependent traffic characteristics for a large-scale traffic network which are reasonable and internally consistent. This is the largest dynamic route choice solution which has been obtained thus far. Conclusions and recommendations for future research are presented as well.

Contents

List of Tables

iv

List of Figures

Preface

The solution of dynamic route choice models for large networks has become an urgent research priority in the context of Intelligent Transportation Systems. This report seeks to contribute to the need for operational models in this field.

The solution algorithm builds on and extends the work of Professor Bruce Janson, University of Colorado at Denver. We are grateful for his enthusiastic collaboration and his detailed comments.

The accident analysis responds to a request from the Illinois Department of Transportation for assistance in modeling the effect of accidents on traffic congestion. This aspect of the research was supported by the ADVANCE Project, a field test conducted by the Illinois DOT and the Federal Highway Administration during 1991-1996. The network data used in the research is from that project.

This report was originally presented by Der-Horng Lee as his Ph.D. thesis in civil engineering. This report is issued in the hopes of stimulating additional comments on the findings.

David E. Boyce

SYMBOLS:

t	a time interval
d	a departure time interval
Δt	duration of each time interval
$\mathcal{N}$	set of all nodes
$\mathcal{Z}$	set of all zones (i.e., trip end nodes)
$\mathcal{A}$	set of all links (i.e., directed arcs)
$\mathcal{T}$	set of all time intervals in the full analysis period (e.g., 18 ten–minute time intervals for a three–hour analysis period)
$x_a[t]$	total flow of vehicles on link a in time interval t
$x_a^{rs}[d]$	flow of vehicles from zone r to zone s on link a that departed in time interval d
$\tau_a(x_a[t])$	actual travel time on link a with flow x_a in time interval t
$\tau_a[t]$	actual travel time on link a in time interval t
π_{ri}^d	minimal actual travel time from zone r to node i for flow departing in interval d
$\bar{\pi}_{ri}^d$	number of time intervals traversed in π_{ri}^d
$\Delta \pi_{ri}^d$	difference of minimal travel times from origin r to node i for flow departing in successive time intervals
$f_{rs}[d]$	departure flow rate at origin r to destination s in departure time interval d
$e_{rs}[t]$	exit flow rate at destination s from origin r in time interval t
$A(j)$	set of links whose tail node is j
$B(j)$	set of links whose head node is j
$q_{rj}[d]$	flow of vehicles from zone r to node j departing in time interval d via any route
$\phi_{ri}^d[t]$	fraction of the flow departing zone r in time interval d that crosses node i in time interval t
Ω_{ra}^d	difference between the minimal travel time from zone r to node i (π_{ri}^d) plus the travel time on link a $(\tau_a[d + \bar{\pi}_{ri}^d];\ a = (i,j))$, and the minimal travel time from zone r to node j (π_{rj}^d), for vehicles departing from zone r in time interval d
$\alpha_{ri}^d[t]$	*node time interval*, a $[0,1]$ indicator of whether the flow departing zone r in time interval d has crossed node i in time interval t
β_{ri}^d	time at which the last flow departing zone r in time interval d crosses node i via its shortest route, less FIFO delay time at node i
$\theta_{ri}^d[t]$	fraction of a time interval t that the last flow departing zone r in time interval d crosses node i
$\mu_{ra}^d[t]$	average travel time on link a of the last flow departing zone r in time interval d

h minimum fraction of time interval separating flows departing in successive time intervals

ψ_a^t fraction of link a occupied by queue during time interval t

ABBREVIATIONS:

ADVANCE	Advanced Driver and Vehicle Advisory Navigation Concept
ATIS	Advanced Traveler Information Systems
ATMS	Advanced Traffic Management Systems
BLPP	Bi-Level Programming Problem
BPR	Bureau of Public Roads
CATS	Chicago Area Transportation Study
CBD	Central Business District
CCTV	Closed-Circuit Television
DUO	Dynamic User-Optimal
FHWA	Federal Highway Administration
FIFO	First-In-First-Out
F-W	Frank-Wolfe
IDOT	Illinois Department of Transportation
ITS	Intelligent Transportation Systems
IVHS	Intelligent Vehicle Highway Systems
K-K-T	Karush-Kuhn-Tucker
NCSA	National Center for Supercomputing Applications
O-D	Origin-Destination
SO	System-Optimal
SUO	Static User-Optimal
TMC	Traffic Management Center
UO	User-Optimal
VI	Variational Inequality

Chapter 1

INTRODUCTION

1.1   Research Background and Problem Statement

Intelligent Transportation Systems (ITS), also known as Intelligent Vehicle Highway Systems (IVHS), make use of advanced technologies (such as navigation, automobile, computer science, telecommunication, electronic engineering, automatic information collection and processing) in an effort to improve the movement of people and goods. ITS technologies have the potential to provide better travel information, easier and safer travel, improved network capacity utilization, less traffic congestion, improved traffic flow, energy consumption savings, quicker roadway emergency response, faster freight deliveries and improved fleet management.

Within the framework of ITS, Advanced Traveler Information Systems (ATIS) can provide historical, real-time and predictive information to support travel decisions, which in turn can influence the travel choices of individuals and consequently improve the time and quality of travel (Ran and Boyce, 1994). Moreover, Advanced Traffic Management Systems (ATMS) can integrate the management of various roadway control functions including ramp metering, signal timing, and variable speed advisories to predict traffic congestion and provide alternative routing instructions to drivers. General roadway information can be broadcast to drivers through AM/FM radio. Conversely, detailed information generated by ATMS can be transmitted to specific groups of drivers via in-vehicle route guidance systems. In consideration of the goals of ATMS and ATIS, dynamic route choice models, based on behavioral assumptions regarding route choice, are needed to perform traffic condition assessment and prediction for generating route guidance. Therefore, dynamic route choice models are an essential element for ATMS and ATIS evaluation and implementation.

The problem analyzed in this research is stated as follows: given time-dependent travel demand (in the form of origin-destination (O-D) matrices for short time intervals), distribute this demand onto routes through the network based on (1) specified route choice behavior

(e.g., user-optimal (UO) or system-optimal (SO)), and (2) time-dependent transportation supply variations. For a specific solution procedure, the interaction between transportation supply and travel demand is determined and represented by the flow pattern of the network.

To date, very few traffic models are adequate for ATMS and ATIS applications. This research seeks to contribute to this need from the following aspects.

1. solve a dynamic route choice model on a realistic, large-scale traffic network;

2. represent the network in detail, including use of realistic traffic engineering-based travel time functions;

3. provide a platform to describe and analyze traffic dynamics in the real world;

4. support and evaluate network-wide traffic signal settings, network-wide traffic control strategies and route guidance.

Route choice models are usually associated with network equilibrium concepts. Static models are generally used for long-term planning purpose. In static models, there are no time-dependent variables within the period of analysis, generally the peak travel period. Most *static user-optimal* (SUO) models are formulated to yield route choices that are consistent with Wardrop's first principle (Wardrop, 1952). This principle requires *route travel costs to be equal for used routes between a given origin-destination pair, with no unused route having a lower cost*. This user-optimal has been employed as the key behavioral assumption in most route choice models.

Conceptually, the model studied in this research seeks to achieve a dynamic generalization of Wardrop's first principle. Therefore, the *dynamic user-optimal* (DUO) problem is defined as determining the route flows at each instant of time that result in drivers using minimal-time routes, and finding the associated link flow pattern.

The problematic assumption that complicates the modeling of dynamic route choices is that route choices must be based on travel times which are temporally-consistent with future link flows. This assumption is appropriate for recurrent trips and traffic conditions, and is also acceptable for *scheduled* events (e.g., ballgames, concerts, parades, detours and road constructions) and even for predicted weather conditions. However, this behavioral assumption is inconsistent with *unexpected* events (e.g., stalled vehicles, dropped objects, dangerous chemical spills and accidents) at future times because drivers have very limited capability to be informed about the times and locations of such events before they encounter unusual queuing delays caused by those incidents. Enroute diversions are thus expected to occur only when incidents are encountered by drivers. To this end, alternative route choice strategies based on anticipatory and non-anticipatory traffic conditions are considered in

this research to determine the range of possible incident impacts from the least severe to the most severe cases.

1.2   Research Objectives

This research aims to present a DUO route choice model for predicting dynamic traffic conditions intended for off-line ATMS and ATIS evaluation and implementation. This DUO route choice model is formulated as a variational inequality (VI) and can be solved efficiently to convergence by the proposed diagonalization algorithm with discrete time intervals. For most links a realistic traffic engineering-based link travel time function, the Akcelik function (Akcelik, 1988), is adopted in this research in place of the simplistic but widely used BPR function (Bureau of Public Roads, 1964) to estimate delays and travel times for various types of links and intersections. Unexpected events, such as stalled vehicles, traffic accidents, dropped objects, and dangerous chemical spills that cause nonrecurrent traffic congestion, are analyzed with the model. Route choice behavior based on anticipatory and non-anticipatory network conditions are considered in performing the incident analysis, extending the capability of this model to contribute to ATMS and ATIS.

The specific objectives of this research can be itemized as follows:

1. formulate a DUO route choice model based on the variational inequality problem;

2. solve the proposed DUO route choice model using a diagonalization algorithm;

3. utilize traffic engineering-based link travel time functions in the solution algorithm;

4. implement the formulated model on a realistic, large-scale traffic network;

5. generate time-dependent traffic information (such as flows, speeds, travel times) classified according to turning movements, highway facilities and traffic control at intersections;

6. understand the solution properties of a DUO route choice model on a realistic traffic network;

7. analyze traffic flow patterns under incident conditions with different route choice assumptions.

1.3   Organization of the Report

This research consists of seven chapters, including this introductory chapter. In Chapter 2, an extensive review of the literature on dynamic route choice modeling is presented. Desired

properties of the dynamic route choice model are introduced, and the desired capabilities of the model to represent traffic dynamics and traveler behavior are discussed. Requirements for theoretical properties and computational issues are identified. A wide spectrum of approaches to dynamic route choice modeling is presented including: optimization-based approach; optimal control theory-based approach; variational inequality-based approach; and simulation-assignment-based approach. The review also includes a general review of variational inequality theory.

Chapter 3 introduces the formulation of the proposed DUO route choice model. The link-time-based conditions are defined. Although the route-time-based VI model has the most straightforward interpretation of DUO route choice (Ran and Boyce, 1994), it requires the enumeration of all routes between each O-D pair. For a continuous time formulation and a network of reasonable scale, route enumeration is intractable (Patriksson, 1994). To apply the model to a large-scale network, therefore, the alternative link-time-based VI model is considered. The dynamic network constraints, such as flow conservation constraints, flow propagation constraints and first-in-first-out (FIFO) constraints, are derived according to the link-time-based conditions. The proposed model aims to find the flow pattern satisfying the defined DUO state. To this end, the travel times experienced by flows for each O-D pair departing at the same instant of time are equal and minimal at the DUO state.

In Chapter 4, the solution algorithm of the proposed VI model is considered. The steps of the algorithm are described in detail. Using the diagonalization method, the proposed VI model can be solved efficiently and smoothly to convergence. Adjustments of link capacities are incorporated into the algorithm to account for capacity changes caused by spillback queuing effects, signal timing changes (if any), incidents and other interruptive events. This function greatly enhances the ability of this model to model dynamic traffic in the real world.

Chapter 5 examines issues of implementing this model on the ADVANCE Network. The ADVANCE Network is located in the northwestern suburbs of Chicago and covers about 300 square miles (770 square kilometers), and has 447 O-D zones, nearly 8,000 links and more than 2,500 nodes. Diversified land use patterns and highway facilities characterize the ADVANCE Test Area. To generate specific link travel times for the investigated network, the expanded intersection representation is employed. Using this network representation, each turning movement is coded as an individual intersection link, increasing the network scale to approximately three times larger than the conventional network representation. Nearly 10,000 nodes and 23,000 links are defined for the solution procedure. Realistic traffic engineering-based link travel time functions, based on the Akcelik function (Akcelik, 1988), are adopted. Each of the links is identified according to type of intersection control, highway facility type, geometric layout and lane designation at intersections to determine the appropriate link travel time function. Daily trip tables from CATS (Chicago Area Transportation

Study) factored to represent travel demand for five time-of-day periods (night, morning peak, mid-day, afternoon peak and evening) are utilized. Each time-of-day period is further divided into ten-minute intervals for use in the solution procedure. Ten-minute trip departure rates for each O-D zone are derived from half-hour departure rates obtained from CATS. Other traffic input data required by the solution procedure are described. Alternative route choice strategies employed by the proposed model to estimate the impacts of unexpected capacity reducing events are presented in detail.

Computational solutions and analyses of results are presented in Chapter 6. First, the computing platform used for solving the proposed model and its performance are described. Then, dynamic network performance measures (average travel time, average travel distance, network space mean speed, average flow-to-capacity ratio and a convergence index) are defined and presented. These network performance measures are analyzed to interpret the overall performance of the proposed model on the ADVANCE Network. Finally, computational results for different locations within the ADVANCE Network under different incident scenarios are analyzed in detail.

In Chapter 7, an account of the major contributions and conclusions of this research are presented. Recommendations for future research are provided as well.

Chapter 2

Literature Review

Recognition of the importance of dynamic network modeling has increased in recent years. Various types of modeling approaches, solution methods and test results have been reported, focusing on dynamic network equilibrium analysis. In general, approaches to the investigation of dynamic network equilibria can be classified into four major categories according to the nature of their methodology: optimization-based approach; optimal control theory-based approach; variational inequality-based approach; and simulation-assignment-based approach.

The principal research contributions in those four methodological categories are reviewed in this chapter. Since the defined dynamic user-optimal (DUO) route choice model is formulated as a variational inequality (VI) problem, this chapter also includes a review of variational inequality theory. Before examining previous works on dynamic network modeling, we first consider the desired properties of dynamic route choice models.

2.1 Desired Properties of Dynamic Route Choice Models

Properties of dynamic route choice models can be addressed from the following viewpoints:

1. the representation of network traffic flow;

2. the delineation of travelers' behavior and characteristics;

3. the existence and uniqueness of the solution;

4. the computational feasibility of the model for real applications;

5. the ability of the model to be integrated into ATMS/ATIS applications.

2.1.1 Representation of Network Traffic Flow

For the representation of dynamic network traffic flow, several basic and critical requirements need to be established. These requirements include representation of flow conservation, FIFO conditions, flow propagation and queue spillbacks over the network. In the static case, these required properties are straightforward or even trivial. They complicate greatly, however, the modeling tasks in the dynamic case.

Conservation of flow on a multiple O-D network needs to be considered both at the level of nodes and links. In addition, flow conservation at origins and destinations must be considered. In a continuous time formulation of the dynamic route choice problem, first-in-first-out (FIFO) conditions are implicitly defined by the flow propagation constraints, if an appropriate link travel time function is applied. In contrast, additional attention needs to be paid to maintain FIFO conditions when a discrete time formulation (or solving a continuous time formulation in a discrete manner) is adopted. For a discrete time formulation, FIFO conditions have to be expressed explicitly to maintain effective trip ordering in successive time intervals. That is, vehicles are assumed to make one-for-one (or zero-sum) exchanges of traffic movements as traversing on any link, which is acceptable and expected in most aggregate traffic models (Janson and Robles, 1995).

Flow propagation is not necessary for static route choice models because each O-D flow propagates instantly over the entire route from origin to destination in the static model. For dynamic route choice models, flows remain on a link only for some duration of time and the movements over time need to be represented. A dynamic route choice model needs to include the capability of capturing this phenomenon.

Representation of queue spillback is another issue closely related to flow propagation. Queue spillbacks are often caused by unexpected capacity reducing events such as accidents, stalled vehicles and dropped objects, so that the traffic demand exceeds the available capacity. Continuing excess traffic demand turns the local oversaturation into areawide oversaturation. Within the context of representing the network traffic, inclusion of the effects of traffic control facilities and strategies, such as traffic signal and ramp metering, are also essential.

2.1.2 Delineation of Travelers' Behavior and Characteristics

Dynamic route choice models must address travelers' route choice behavior and characteristics. One basic hypothesis used in this research is the DUO concept. DUO behavior implies that travelers choose the best routes based on either *instantaneous* or *actual* travel times. *Instantaneous* travel time is defined as the travel time based on currently prevailing traffic conditions. *Actual* travel time is defined as the travel time actually experienced during the

trip. Consequently, the *instantaneous* DUO state and the *ideal* DUO state are defined.

The *instantaneous* DUO route choice problem is to determine flows at each instant of time on each link resulting from travelers using minimal-time routes under currently prevailing travel times. The corresponding model provides currently prevailing traffic information to travelers. However, *instantaneous* route flows with the same departure time and the same O-D may actually experience different route travel times, because the route time may subsequently change due to rapidly changing traffic conditions, even though at each *decision node* the flows select the route that is currently best. A *decision node* for each route of each O-D pair is defined as any node on the route including the origin where an O-D flow can switch to an alternative route toward its destination. This *instantaneous* DUO definition is also referred as the *reactive* DUO state (Jayakrishnan et al., 1995), since travelers seek to minimize their own travel times by continuously updating their route choices according to currently prevailing traffic conditions. Using the *instantaneous* DUO route choice model can lead to inferior solutions and/or unrealistic predictions of traffic patterns because travelers choose routes without anticipating future traffic conditions.

For the *ideal* DUO route choice problem, for each O-D pair at each instant of time, the *actual* travel times experienced by travelers departing at the same time are equal and minimal. The determined traffic flow pattern over the network is called a travel-time-based *ideal* DUO state. This *ideal* DUO definition is also referred as the *predictive* DUO state (Ran and Boyce, 1994) since the *actual* route travel time is predicted using the corresponding route choice model. Under the *ideal* DUO state, travelers have no reason to change their routes.

In addition, deployable dynamic route choice models should use generalized cost functions by weighting travel times, operating costs and other associated variables instead of travel times only. To represent fully travelers' route choice behavior and characteristics, multiple classes of travelers (e.g., ATIS equipped, degree of ATIS compliance, knowledge of network), multiple classes of drivers (e.g., aggressive, conservative) and multiple classes of vehicles (e.g., car, bus, truck) need to be introduced in future models.

2.1.3   Existence and Uniqueness of the Solution

Dynamic route choice models should have a solid theoretical foundation to ensure the existence and uniqueness of the DUO solution. These properties pertain to both simulation and analytical approaches. In principle, simulation-assignment-based approach lacks proofs of existence and uniqueness of the equilibrium solution, as well as convergence of the algorithm to that solution. Therefore, flow patterns generated by simulation-assignment-based models are difficult to compare for alternative scenarios.

Analytical-based models, such as optimization and variational inequality models, have

advantages in this regard. Moreover, for a dynamic route choice model, any claim for DUO routes must be proven because the time-dependent shortest route approach does not necessarily lead to DUO routes under congestion.

2.1.4 Computational Feasibility of the Model for Realistic Applications

Computational feasibility is one of the most challenging issues of dynamic route choice modeling. Since the first large-scale solution of an analytical-based dynamic route choice model has obtained (Boyce et al., 1995a), however, solution of dynamic route choice models to large networks can no longer be regarded as an infeasible task. Supercomputers are not the only possible platform for dynamic route choice models. Workstations can be an appropriate platform in considering the design of a Traffic Management Center (TMC). High performance computing techniques, such as distributed and parallel computation, are worth attempting in pursuing savings in computational times. In addition, a balance between computational speed and accuracy of solution should be maintained.

2.1.5 Capabilities for ATMS/ATIS Applications

The capabilities of a dynamic route choice model for ATMS/ATIS applications can be analyzed from two viewpoints: real-time application and off-line evaluation. In principle, dynamic route choice models are designed for real-time application; therefore, acquisition and assimilation of real-time traffic information become essential. The fusion of results from dynamic route choice models with other relevant simulation and statistical models is critical for providing dynamic information of flows, queues and travel times. A well-organized system architecture/interface between a dynamic route choice model and various traffic information sources (e.g., probes and surveillance systems) and compatibility with various TMC architectures (e.g., centralized, decentralized, distributed) are required for deployment.

Dynamic route choice models can be used for off-line evaluations of ATMS/ATIS and common traffic control strategies such as signal control, ramp metering and reversible lane allocation. Based on dynamic route choice models, off-line generated routes can be used for establishing the route data base for a route guidance system. Likewise, off-line generated link information can be used in the data fusion process to provide further estimates of traffic characteristics. Off-line applications of dynamic route choice models also extend to environmental impact analysis and forecasting for urban transportation planning.

2.2 Approaches to Dynamic Route Choice Modeling

According to the methodology adopted, approaches used to investigate the dynamic route choice problem can be generally classified into four major categories: optimization; optimal control theory; variational inequality; and computer simulation. Representative literature from these four approaches is reviewed in the following subsections.

2.2.1 Optimization-Based Approach

The pioneering effort in this field was by Merchant and Nemhauser (1978a, 1978b). Their model (M-N) is formulated as a discrete time, nonconvex and nonlinear programming problem of system-optimal (SO) route choice to a single destination. A toy network with multiple origins and a single destination was used for generating numerical results. Using the one-pass simplex method, the M-N model was solved to a global optimum with a piecewise linearization of the objective function of the model (Merchant and Nemhauser, 1978a). After the linearization of objective function, the M-N model exhibits a perfect mathematical staircase structure so that the linear decomposition technique for sparse matrices can be applied to solve the transformed linear program.

The K-K-T (Karush-Kuhn-Tucker) optimality conditions show that a dynamic generalization of Wardrop's second principle (SO; Wardrop, 1952), requiring equal marginal travel costs for used routes, is obtained (Merchant and Nemhauser, 1978b). The M-N model was also examined under steady-state assumptions; and the model was proven to be a proper generalization of the conventional *static system-optimal* route choice model. From today's viewpoint, the greatest contribution of the M-N model is that traffic congestion was treated explicitly in the flow constraints.

Several numerical efforts and extensions have been made based on the M-N model. Ho (1980) showed that the global optimum of the M-N model could be obtained by solving a sequence of at most $N + 1$ linear programs, where N is the number of time periods. Carey (1986) re-solved the M-N model and showed that it satisfies a linear independence constraint qualification which establishes the validity of the optimality analysis presented by Merchant and Nemhauser (1978b). Later, Carey (1987) reformulated the M-N model for multiple destinations as a convex, nonlinear program with nonlinear constraints for each link. This new formulation of the model has analytical, computational and interpretation advantages over the original formulation (1978a). In particular, the K-K-T conditions are both necessary and sufficient to characterize an optimal solution of Carey's reformulation, whereas in the original model the K-K-T conditions are not sufficient because the constraint set is nonconvex.

Building on the M-N model, Ho (1990) proposed an extended stochastic formulation of

the dynamic network with congestion by relaxing the assumption that exogenous flows into nodes are known for all time periods; thus, there are uncertain input flows to each link in this model. A successive optimization procedure, called the nested decomposition algorithm similar to the algorithm used by Ho (1980), was adopted to obtain a globally optimal solution and convergence. Some computational results are generated by using Ho's algorithm. These models and numerical examples, however, were tested only on small, hypothetical networks.

Janson (1991a) presented the first attempt of solving a mathematical programming formulation of a DUO route choice model on a real network. This DUO model was formulated as a nonlinear, mixed-integer program in terms of route-flow variables with scheduled departure times and variable arrival times. His paper described a dynamic route choice heuristic that generates approximate solutions to the DUO state for large networks. Janson (1991b) proposed a link-flow formulation of the DUO route choice problem and a convergent solution algorithm. Both of the route-flow and link-flow formulations presented by Janson (1991a, 1991b) are indeed a temporal generalization of static user-optimal (SUO) route choice problem for a multiple O-D network with additional constraints to ensure temporal continuous flow propagation and route flows. In Janson's model, the DUO state is defined as (Janson, 1991b):

1. All routes between a given pair of zones used by trips departing in a given time interval must have equal travel costs.

2. No route between a given pair of zones not used by trips departing in a given time interval cannot have a lower travel cost.

Hence, his model is an example of an *ideal* DUO. Janson and Robles (1993) presented a DUO formulation with arrival time costs. That model was formulated as a bi-level programming problem (BLPP) consisting of an upper and lower problem and solved successively by an iterative algorithm claimed to converge to satisfy the necessary optimality conditions of the problem. The above models of Janson are all formulated in discrete time.

Most recently, Janson and Robles (1995) converted the discrete time modeling approach into a quasi-continuous time formulation for their previous models. Three key model improvements are obtained using the new approach.

1. Traffic flows are spread over time intervals in continuous time which allow trips to be split among successive time intervals.

2. FIFO trip ordering between all O-D zones is more precisely maintained.

3. The performance of flow propagation and queue spillback estimation is improved.

11

The new formulation maintains the BLPP framework. The upper-level problem solves a multi-interval, time-varying-demand route choice problem. The lower-level problem maintains temporally-correct, time-continuous traffic flow propagations. Their algorithm (Janson, 1991b; Janson and Robles, 1995) is applied to solve these two subproblems to convergence. Besides the modeling refinements, computational results are presented for the I–25/HOV corridor (110 zones, 1,714 nodes, 3,417 links and 222,218 O-D trips) located in southeast Denver (Robles and Janson, 1995). The BPR function is used in this model. The solution is validated with data obtained from in-pavement detectors. The entire dynamic route choice framework of Janson is called DYMOD for easy reference.

De Romph (1994) modified an early version of DYMOD called 3–DAS (3–Dimensional Assignment) to the Washington, D.C. network (180 zones, 857 nodes and 2,086 links) and the Amsterdam network (21 zones, 286 nodes and 430 links) for ATMS applications. A speed-density function proposed by Smulders (1988) was used for calculating travel times. A graphical user interface of 3–DAS was also established to demonstrate results.

Jayakrishnan et al. (1995) applied a modification of DYMOD to a hypothetical 5 × 5 network and a small-scale network (38 zones, 416 nodes, 914 links and 1,406 O-D trips) located in Anaheim, California. Modified Greenshields speed-density relationships were used to derive a link-cost function that is monotonically nondecreasing and convex with respect to density. The order of the upper-level and lower-level problems of DYMOD was reversed by Jayakrishnan et al. (1995), which led to a discussion of the appropriateness of interpreting DYMOD as a bi-level programming problem and a Stackelberg leader-follower game (von Stackelberg, 1952).

Boyce, Lee, Janson and Berka (1995a, 1995b, 1996) incorporated realistic traffic engineering-based link travel time functions into the solution algorithm of DYMOD to estimate better the link travel times and intersection delays. The modification was implemented for the AD-VANCE Network (Boyce et al., 1994) (447 zones, 9,700 nodes, 23,000 links) located in the northwest Chicago area. A link-time-based variational inequality formulation of DYMOD was proposed by Boyce, Lee, Janson and Berka (1995b) to provide an improved theoretical basis for the solution algorithm. Dynamic network constraints such as flow conservation constraints, flow propagation constraints and FIFO constraints were developed. Using alternative route choice strategies, unexpected capacity reducing events that caused nonrecurrent traffic congestion were also analyzed by the model. Network performance measures were defined and computational results were obtained and compared with a large-scale asymmetric static route choice model (Berka et al., 1994). The results of Boyce, Lee, Janson and Berka (1995a, 1995b, 1996) established a new benchmark for solving a dynamic route choice model for a large-scale network.

2.2.2 Optimal Control Theory-Based Approach

Optimal control theory is suitable for describing and optimizing time-dependent dynamic processes. Luque and Freisz (1980) proposed the first dynamic route choice model that applied optimal control theory. They reformulated the M-N model (Merchant and Nemhauser, 1978a) as a continuous-time optimal control problem. The optimality conditions were derived from Pontryagin's minimum principle (Pontryagin et al., 1962) and can be interpreted as the dynamic generalization of Wardrop's second (SO) principle for the static case. Friesz et al. (1989) offered a DUO route choice model by considering the equilibration of instantaneous route costs.

Subsequently, a generalized DUO route choice model of a multiple O-D network was presented by Wie et al. (1990). Wie (1991) analyzed a simple dynamic extension of the static route choice problem with elastic demand. The problem of DUO route choice with elastic demand is not only to predict the dynamic traffic pattern but also to determine the temporal distribution of traffic from each origin in response to time-dependent traffic conditions. Using the augmented Lagrangian method, Wie et al. (1994) solved dynamic route choice models in discrete time. In the proposed algorithm, the need for route enumeration was obviated. The algorithm also exploited the natural decomposition of the route choice problem by time period which is possible when an optimal control formulation is employed.

In the formulation of some optimal control-based dynamic route choice models (Friesz et al., 1989; Wie 1989; Ran and Shimazaki, 1989a), only the inflow into each link at a given instant of time is defined as a control variable; the exit flow from each link is considered to be a function of the number of vehicles on that link. Although this mechanism provides an explicit relationship between exit flow and number of vehicles, it has several critical drawbacks, as follows.

1. If the exit flow function is concave, it is impossible to establish an optimal control model of the DUO route choice problem for a multiple O-D network.

2. If the initial link inflow is zero, flow propagation tends to be unrealistic when the inflow comes positive, since the exit flow rate must be positive immediately to satisfy the exit flow function.

By defining exit flow as a control variable, Ran and Shimazaki (1989b) proposed a DUO route choice model which considers the equilibration of instantaneous travel times. Its computational complexity is significantly reduced compared with only defining the inflow as a control variable. Moreover, the resulting model can be applied to a multiple O-D network. Further, Ran, Boyce and LeBlanc (1993) formulated a new class of *instantaneous* DUO route choice models with flow propagation constraints which generalized the SUO route

choice model. Since these formulated optimal control problems are convex programs with linear constraints, unique solutions are obtained. Boyce, Ran and LeBlanc (1995) presented an algorithm for solving their *instantaneous* DUO route choice model. By using an expanded time-space network, the Frank-Wolfe linear programming subproblem only requires the solution of minimal-cost route problems for each O-D pair. Thus, this expansion technique allows standard static route choice algorithms to solve dynamic route choice problems. Ran and Boyce (1994) collected their findings into a book on dynamic transportation network modeling. Complete and rich sets of dynamic route choice models, constraints, optimality conditions, solution algorithms and computational examples are presented.

2.2.3 Variational Inequality-Based Approach

Compared to optimization and optimal control approaches, the variational inequality (VI) approach provides more general formulations of dynamic route choice problems (Ran and Boyce, 1994). The first statement of a network equilibrium in the form of a variational inequality was the SUO route choice model of Smith (1979). Dafermos (1980) developed an elastic demand model with disutility functions using the variational inequality approach. An elastic demand model with demand functions was introduced by Dafermos and Nagurney (1984). Fisk and Boyce (1983) also presented a set of alternative VI formulations for network equilibrium travel choice problems. Nagurney (1993) summarized the modeling and algorithmic aspects of VI models for static route choice problems.

Using the variational inequality approach, Friesz et al. (1993) formulated a simultaneous departure time/route choice model. Smith (1993) proposed a route-based VI formulation using the packet representation of vehicle groups. Wie et al. (1995) formulated the DUO route choice problem as a variational inequality problem in discrete time in terms of route cost functions. A heuristic algorithm is employed to generate numerical results on a hypothetical network (four arcs, two origins and one destination) and the well-known Sioux Falls, South Dakota test network.

Inevitably, explicit route enumeration is required by solution procedures of route-based VI models. However, since dynamic traffic flow does not have constant flow rate during flow propagation, the route-based VI can not be transformed into a link-based VI. Therefore, it is very difficult to develop a solution algorithm for a route-based VI without explicit route enumeration. Explicit route enumeration for large-scale networks with dynamic flows is intractable, especially for continuous time formulations. This characteristic makes the route-based VI models impossible for realistic applications.

To overcome the critical drawback of route-based VI dynamic route choice models, Ran and Boyce (1995) presented a link-based VI DUO route choice model. In addition to devel-

oping the link-based formulation, this model has a traffic engineering orientation, which is more appropriate for realistic applications. Based on this link-based VI model, Ran, Hall and Boyce (1995) developed a link-based VI model for the dynamic departure time/route choice problem. Boyce, Lee, Janson and Berka (1995b) proposed a link-time-based VI formulation to provide an improved theoretical basis for the solution algorithm of DYMOD, which was applied to the ADVANCE Network.

2.2.4 Simulation-Assignment-Based Approach

In reviewing the simulation-assignment-based approach to the dynamic route choice problem, we use the term *simulation-assignment* instead of *simulation* to distinguish this approach from simulation models that do not consider route choice. This group of models adopts car-following simulation techniques, which relate speed and density on the link to macroscopic relations, but move vehicles individually or in platoons. This approach is situated somewhere between microscopic simulation and macroscopic simulation, and is sometimes referred as *mesoscopic* (De Romph, 1994).

Two models using this approach are INTEGRATION (Van Aerde and Yagar, 1988) and DYNASMART (Mahmassani and Peeta, 1993). Both models are designed in the context of in-vehicle route guidance systems. Various speed-flow functions are available in INTEGRATION which are determined by using four different user-defined parameters (free flow speed, speed at capacity, capacity and jam density). Routes are based on real-time traffic conditions and are instantaneous. Intersection delays, traffic signal effects and different classes of drivers are able to be modeled by INTEGRATION. DYNASMART simulates individual vehicles, moving them at speeds determined by the total flow on the link. A modified Greenshields speed-density function is adopted for DYNASMART. The selection of a route is decided by each individual vehicle at each decision point in the network. Intersection delays are explicitly modeled within DYNASMART. Multiple classes of drivers are defined in DYNASMART which facilitates its use in simulating various ATIS scenarios.

2.2.5 Notes

Optimization-based approach has a long standing history in dynamic route choice modeling. Many solution algorithms have been developed and applied to optimization-based dynamic route choice models (e.g., Janson and Robles, 1995; Boyce, Lee, Janson and Berka, 1995a, 1996). Although more theoretical limitations are found in optimization-based models, which may cause these models to fail to provide good descriptions of traffic dynamics in certain situations, optimization-based models are feasible to solve, as compared with optimal control-based models.

Optimal control theory is suitable for describing a time-dependent dynamic process. Optimal control-based dynamic route choice models exhibit sound theoretical formulations and mathematical derivations, and provide continuous-time formulations which correspond to the defined equilibrium states. However, computationally practical procedures for solving large-scale control theory-based route choice models are not available.

Variational inequality-based models provide more generalized formulations of dynamic route choice problems. Using the link-based VI formulation, the DUO route choice model can be solved on a large-scale network (Boyce, Lee, Janson and Berka, 1995b).

Compared with optimization and optimal control approaches, simulation-assignment models lack an analytical model formulation. Proofs of existence, uniqueness and convergence of solutions are unavailable for simulation-assignment models. Simulation-assignment models, however, are easier to be implemented than analytical-based dynamic route choice models for traffic control schemes applied in ATMS/ATIS because there is no concern of violating the solution properties which is crucial for an analytical-based model. In addition, for an analytical-based dynamic route choice model, a complicating factor in modeling many route choice options is that route choice decisions must be based on travel times which are temporally-consistent with future link flows. This restriction makes the incorporation of traffic controls schemes applied in ATMS/ATIS into analytical-based dynamic route choice models more challenging.

2.3 Review of Variational Inequality Theory

Since our DUO route choice model is formulated as a variational inequality, the review of literature includes an overview of variational inequality theory. The variational inequality problem is a general formulation that encompasses a set of mathematical problems, including nonlinear equations, optimization problems, complementarity problems and fixed point problems. Variational inequalities were originally developed as a tool for the study of certain classes of partial differential equations such as those that arise in mechanics. This section is based on Nagurney (1993) and Ran and Boyce (1994).

2.3.1 Definitions for Variational Inequality Problems

In this section, we present several types of variational inequality problems. First, we discuss the variational inequality for static problems. Here, we are concerned with a vector of decision variables $x = (x_1, x_2, \cdots, x_n)$ and a vector of cost functions $f(x) = [f_1(x), f_2(x), \cdots, f_n(x)]$. Define G as a given closed convex set of the decision variables x; f is a vector of given continuous functions defined on $\Re^n$. Then, we define the static case as follows.

Definition 2.3.1. *The finite-dimensional variational inequality problem is to determine a vector $x^* \in G \subset \Re^n$, such that*

$$f[x^*] \cdot [x - x^*] \geq 0 \qquad \forall\, x \in G \qquad (2.1)$$

In geometric terms, variational inequality (2.1) states that $f(x^*)$ is orthogonal to the feasible set G at the point x^*.

Now, we define the variational inequality problem for dynamic models. The continuous time formulation is presented first followed by its transformation to discrete time problems for consistency with the subsequently proposed DUO route choice model. First, consider a vector of control variables $u(t) = [u_1(t), u_2(t), \cdots, u_m(t)]$ and their dynamic processes

$$\dot{x}(t) = h[x(t), u(t)]$$

with state variables $x(t) = [x_1(t), x_2(t), \cdots, x_n(t)]$ and state equations $h = [h_1(t), h_2(t), \cdots, h_n(t)]$. Associated with the dynamic processes, there is a vector of cost functions $F(t) = [F_1(t), F_2(t), \cdots, F_m(t)]$. Each element of the cost function vector is a function of the state and control variables:

$$F_i(t) = F_i[x(t), u(t)] \qquad i = 1, 2, \cdots, m$$

Since the state variables $x(t)$ can be determined by the state equations when the control variables $u(t)$ are given, the vector of cost functions can be further simplified as $F(t) = F[u(t)]$. Let $G(t)$ be a given closed convex set of the control variables $u(t)$. We assume $F(t)$ is a set of given continuous functions from $G(t)$ to $\Re^n(t)$. Then, we give the following definition of the dynamic variational inequality problem.

Definition 2.3.2. *The finite-dimensional variational inequality problem is to determine a control vector $u^*(t) \in G(t) \subset \Re^n(t)$, such that*

$$\int_0^T F^T[u^*(t)] \cdot [u(t) - u^*(t)]\, dt \ \geq 0 \qquad \forall\, u(t) \in G(t) \qquad (2.2)$$

Many dynamic transportation network equilibrium problems can be formulated as systems of equations. The systems of equations can be written as

$$F[u^*(t), x^*(t)] = 0 \qquad (2.3)$$

This problem can also be regarded as a special case of a variational inequality.

2.3.2 Conditions for Existence and Uniqueness

Next we discuss the existence and uniqueness of the solution of the variational inequality problem. For concise notation, conditions for static problems are provided here. However,

the conclusions also apply to variational inequalities for the dynamic problems. Existence of a solution to a variational inequality problem follows from continuity of the function f entering the variational inequality, provided that the feasible set G is compact. In general, we have the following existence theorem.

Theorem 2.3.1. *If G is a compact convex set and $f(x)$ is continuous on G, then the variational inequality problem has at least one solution x^*.*

Proving this theorem requires the use of Brouwer's Fixed Point Theorem (Nagurney, 1993).

Qualitative properties of existence and uniqueness are obtained under certain monotonicity conditions. First, we present the following definitions.

Definition 2.3.3. *A vector of functions $f(x)$ is monotone on G if*

$$[f(x^1) - f(x^2)] \cdot (x^1 - x^2) \geq 0 \qquad \forall\, x^1, x^2 \in G \tag{2.4}$$

where x^1 and x^2 are any two points on G.

Definition 2.3.4. *A vector of functions $f(x)$ is strictly monotone on G if*

$$[f(x^1) - f(x^2)] \cdot (x^1 - x^2) > 0 \quad \forall\, x^1, x^2 \in G;\ x^1 \neq x^2 \tag{2.5}$$

Definition 2.3.5. *A vector of functions $f(x)$ is strongly monotone on G if for some $\alpha > 0$*

$$[f(x^1) - f(x^2)]^T \cdot (x^1 - x^2) \geq \alpha\, ||x^1 - x^2||^2 \quad \forall\, x^1, x^2 \in G \tag{2.6}$$

Assume that $f(x)$ is continuously differentiable on G and $\nabla f(x)$ is strongly positive definite. Then $f(x)$ is strongly monotone. Then, we have the following theorem for uniqueness.

Theorem 2.3.2. *Suppose that $f(x)$ is strictly monotone on G. Then, the solution is unique, if one exists.*

In the following, we present some methods for checking the monotonicity of functions.

Theorem 2.3.3. *Suppose that $f(x)$ is continuously differentiable on G and the Jacobian matrix*

$$\nabla f(x) = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\[2mm] \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} & \cdots & \dfrac{\partial f_2}{\partial x_n} \\[2mm] \vdots & \vdots & \ddots & \vdots \\[2mm] \dfrac{\partial f_n}{\partial x_1} & \dfrac{\partial f_n}{\partial x_2} & \cdots & \dfrac{\partial f_n}{\partial x_n} \end{bmatrix}$$

is positive semidefinite (or positive definite), then $f(x)$ is monotone (or strictly monotone).

Theorem 2.3.4. *Assume that f is continuously differentiable at some $\bar{x}$. Then $f(x)$ is locally strictly (or strongly) monotone at $\bar{x}$ if $\nabla f(\bar{x})$ is positive definite (or strongly positive definite), that is,*

$$v^T f(\bar{x})v > 0 \qquad \forall\, v \in \Re^n;\ v \neq 0 \tag{2.7}$$

$$v^T \nabla f(\bar{x})v \geq \alpha ||v||^2 \quad \text{for some} \quad \alpha > 0, \quad \forall\, v \in \Re^n \tag{2.8}$$

where v is an arbitrary vector with components of real values.

Given these two theorems for monotonicity, we have the following theorem for uniqueness.

Theorem 2.3.5. *Assume that $f(x)$ is continuously differentiable on G and that $\nabla f(x)$ is strongly positive definite, then $f(x)$ is strongly monotone.*

The following theorem provides a condition under which both existence and uniqueness of the solution to the variational inequality problem are guaranteed. No assumption on the compactness of the feasible set G is made, which is important for very complicated dynamic problems when convexity of the feasible set is difficult to prove.

Theorem 2.3.6. *If $f(x)$ is strongly monotone, then there exists precisely one solution x^* to the variational inequality.*

The proof of existence follows from the fact that strong monotonicity implies coercivity, whereas uniqueness follows from the fact that strong monotonicity implies strict monotonicity. In conclusion, in the case of an unbounded feasible set G, strong monotonicity of the function f guarantees both existence and uniqueness. If G is compact, then existence is guaranteed if f is continuous, and only the strict monotonicity condition is needed for uniqueness to be guaranteed. The first conclusion is important for some complicated dynamic problems.

Chapter 3

A Variational Inequality Model of the Dynamic User-Optimal Route Choice Problem

In this chapter, we present a link-based variational inequality (VI) model for the dynamic user-optimal (DUO) route choice problem. VI models provide the most generalized formulation for describing a dynamic network equilibrium. Although route-based VI models have an intuitive interpretation, their computational complexity makes them intractable for realistic applications because of the route enumeration requirement. Dynamic network constraints, link-time-based conditions and the model formulation are described in the following sections. First, we provide a conceptual framework for the *instantaneous* DUO route choice problem and the *ideal* DUO route choice problem.

3.1 Instantaneous vs. Ideal DUO State

An important dynamic generalization of the static user-optimal (SUO) concept is DUO route choice. DUO behavior can imply that travelers choose their best routes based either on *instantaneous* or *actual* travel times. Recall the definitions of *instantaneous* and *actual* travel times in Section 2.1.2. *Instantaneous* travel time is defined as the travel time based on the currently prevailing traffic conditions. *Actual* travel time is defined as the travel time actually experienced during the trip. Consequently, the *instantaneous* DUO state and the *ideal* DUO state are determined.

3.1.1 Instantaneous DUO State

The *instantaneous* DUO route choice problem is to determine vehicle flows at each instant of time on each link resulting from drivers using minimal-time routes under currently prevailing

travel times. The link-time-based *instantaneous* DUO state is defined as (Ran and Boyce, 1994):

> **Link-Time-Based Instantaneous DUO State:** *If, for each O-D pair at each decision node at each instant of time, the instantaneous travel times to the destination over all routes that are being used equal the minimal instantaneous route travel time, the dynamic traffic flow over the network is in a link-time-based instantaneous dynamic user-optimal state.*

Although *instantaneous* user-optimal travel times for all routes that are being used are equal at each decision node at each instant of time, route flows with the same departure time and the same O-D may actually experience somewhat different route travel times. This is because the route travel time may subsequently change due to rapidly changing traffic conditions, even though at each decision node the flows select the route that is currently the best. This *instantaneous* DUO definition is also referred as the *reactive* DUO state (Jayakrishnan et al., 1995), since travelers seek to minimize their own travel times by continuously updating the route choices according to currently prevailing traffic conditions.

3.1.2   Ideal DUO State

An alternative definition of DUO is called *ideal* DUO. The *ideal* DUO route choice problem is to determine vehicle flows at each instant of time so that the *actual* travel times experienced by vehicles departing at the same time and with the same O-D attributes are minimal and equal. The travel-time-based *ideal* DUO state is defined as (Ran and Boyce, 1994):

> **Travel-Time-Based Ideal DUO State:** *For each O-D pair at each interval of time, if the actual travel times experienced by travelers departing at the same time are equal and minimal, the dynamic traffic flow over the network is in a travel-time-based ideal dynamic user-optimal state.*

This *ideal* DUO definition is also referred as the *predictive* DUO state (Ran and Boyce, 1994), since the *actual* route travel time is predicted using the corresponding route choice model. Under the *ideal* DUO state, travelers have no reason to change their routes. Therefore, the obtained DUO state can be viewed as an equilibrium.

3.2   Dynamic Network Constraints

The dynamic network constraints of the proposed link-time-based VI model are presented in this section. The constraint sets include definitional constraints, nonnegativity constraints,

flow conservation constraints, flow propagation constraints and FIFO constraints. We consider a multiple O-D network that is represented by a directed graph $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ where $\mathcal{N}$ is the set of nodes and $\mathcal{A}$ is the set of directed links. Note a node can represent an origin or a destination, as well as an intersection. In the following subsections, the index r denotes an origin and the index s denotes a destination. Both d and t ($t \geq d$) denote a time interval. However, d denotes the departure time interval and t denotes a specific time interval during the journey. Platoons are defined as the traffic flows departing from origin zones in successive time intervals. The constraints described in this section that convert a discrete time formulation into a quasi-continuous time formulation are based on Janson and Robles (1995).

3.2.1 Definitional Constraints

Consider a fixed time period $[0, \mathcal{T}]$ which is long enough to allow all vehicle flows departing during the peak period to complete their journeys. Let

$x_a[t]$ $=$ total flow on link a in time interval t;

$x_a^{rs}[d]$ $=$ flow of vehicles from origin r to destination s on link a that departed in time interval d.

Therefore, for total flow on link a in time interval t, Equation (3.1) must hold.

$$x_a[t] = \sum_{d=1}^{t} \sum_{rs} x_a^{rs}[d] \; \phi_{ri}^d[t] \qquad \forall \, a, t; \; a = (i, j) \tag{3.1}$$

where $\phi_{ri}^d[t]$ is the fraction of all flows departing zone r in time interval d that crosses node i in time interval t. Equation (3.1) defines total flow on link a in time interval t to be the sum of flows departing from zone r in any time interval d from interval 1 up to and including t ($t \geq d$) using link a in time interval t. For a non-continuous time formulation, $\phi_{ri}^d[t]$ is needed to load the flows onto the network and maintains temporally continuous routes in successive time intervals.

3.2.2 Flow Conservation Constraints

For a dynamic route choice model, flow conservation needs to be discussed for different types of nodes including intermediate nodes, origins and destinations. π_{ri}^d is defined as the minimal travel time actually experienced by flows departing from origin r to node i in time interval d, where $\bar{\pi}_{ri}^d$ denotes the number of time intervals traversed in π_{ri}^d and Δt is defined as the duration of each time interval.

$$\bar{\pi}_{ri}^d = \omega \quad \text{if} \quad \omega \leq \pi_{ri}^d / \Delta t < \omega + 1 \tag{3.2}$$

where ω is an integer ($0 \leq \omega \leq \mathcal{T}$). Equation (3.2) makes the actual travel time π_{ri}^d equal to a multiple of the time increment Δt. Define $q_{rj}[d]$ as flows from zone r to node j departing in time interval d via any route. Let $x_a^{rsp}[t]$ be the flow on link a in time t from zone r to s on route p. Equation (3.3) constrains the inflow minus the outflow at any intermediate node j ($j \neq r, s$) in each time interval ($t \geq d$) to the proper departure flows in each time interval between each O-D pair.

$$\sum_{d=1}^{t} \phi_{rj}^d[t] \, q_{rj}[d] = \sum_{sp} \sum_{d=1}^{t} \left[\sum_{a \in B(j)} x_a^{rsp}[d] \, \phi_{ri}^d[t] - \sum_{a' \in A(j)} x_{a'}^{rsp}[d] \, \phi_{rj}^d[t] \right] \tag{3.3}$$

$$\forall \, r, j, t; \; a = (i, j); \; a' = (j, k); \; p$$

$$\text{where} \quad \sum_p q_{rj}^p[d] = q_{rj}[d] \quad \forall \, r, j, d$$

$A(j)$ is the set of links whose tail node is j, and $B(j)$ is the set of links whose head node is j.

Conservation of flow at origin r requires the flow originating at node r in time interval d to equal the flow entering the links leaving origin r in time interval d. Equation (3.4) states the flow conservation at origins.

$$f_{rs}[d] = \sum_{a \in A(r)} \sum_p x_a^{rsp}[d] \qquad \forall \, r, s, d \tag{3.4}$$

Similarly, conservation of flow at destination s requires the flow exiting at node s in time interval t to equal the flow entering destination s in time interval t. The flow conservation at destinations is expressed by Equation (3.5).

$$\sum_{a \in B(s)} \sum_p x_a^{rsp}[t] = e_{rs}[t] \qquad \forall \, r, s, t \tag{3.5}$$

Note that $e_{rs}[t]$ is a variable; Equation (3.5) describes the solution of the model, but does not constrain $x_a^{rsp}[t]$ in this version.

3.2.3   Flow Propagation Constraints

For static route choice models, flow propagation constraints are not necessary because the flow from origin to destination propagates instantly over the entire route. However, in a dynamic route choice model, flow will remain on a link for a duration of time; therefore, the representation of time-dependent flow propagation needs to be considered.

The proposed DUO route choice model requires nonlinear mixed-integer constraints with *node time intervals* ($\alpha_{ri}^d[t]$) and *flow fractions* ($\phi_{ri}^d[t]$) indicating the time intervals where flows originating from each origin cross each node in order to maintain temporally-correct routes

and time-continuous flow propagations over time intervals. Recall in Section 3.2.1, $\phi_{ri}^d[t]$ is defined as the fraction of all flows departing zone r in time interval d that crosses node i in time interval t. Define $\alpha_{ri}^d[t]$ as a [0,1] variable indicating whether the flow departing zone r in time interval d has crossed node i in time interval t. Each *node time interval* acts as an (*if-then*) operator to activate or deactivate certain constraints. A *node time interval* only applies to the last trip (strictly, the end of the platoon or pulse of flow) departing in each departure time interval. The difference between the crossing times at node i of the flows departing in successive time intervals, defined as $\Delta\pi_{ri}^d$, is applied to equations (3.6)–(3.8) to determine the temporal spread of trips crossing node i from the same origin.

$$\phi_{ri}^d[t-k] = \left\{ \min\left[1, \left(\pi_{ri}^d - (t-1)\Delta t\right)/\Delta\pi_{ri}^d\right]\right\}\alpha_{ri}^d[t] \qquad \forall\, k = 0 \tag{3.6}$$

$$\phi_{ri}^d[t-k] = \left(\Delta t/\Delta\pi_{ri}^d\right)\alpha_{ri}^d[t] \quad \forall\, k > 0 \ \text{ for which } \ \pi_{ri}^{d-1} - (t-1-k)\Delta t \le 0 \tag{3.7}$$

$$\phi_{ri}^d[t-k] = \left\{ \max\left[0, \left(\Delta t(t-k) - \pi_{ri}^{d-1}\right)/\Delta\pi_{ri}^d\right]\right\}\alpha_{ri}^d[t] \tag{3.8}$$

$$\text{for minimal } k \text{ for which } \pi_{ri}^{d-1}(t-1-k)\Delta t > 0$$

$$\text{where} \quad \Delta\pi_{ri}^d = \pi_{ri}^d - \pi_{ri}^{d-1} \ \text{ and } \ \pi_{ri}^0 = \pi_{ri}^1 - \Delta t \ \ \forall\, r, i, d$$

where k is used to count the number of boundaries of time intervals spanned by the difference in node crossing times of the last vehicle in the platoon between successive time intervals.

We now use Figure 3.1 to explain Equations (3.6)–(3.8). In Figure 3.1, the y-axis denotes the departure time intervals of platoons and the x-axis denotes the sequence of node along an example route. As shown in Figure 3.1, platoon 1 departs a given origin in its departure time interval 1; the last vehicle of platoon 1 crosses node A in time interval 3. Similarly, Platoon 2 departs the same origin as platoon 1 in time interval 2 and the last vehicle of platoon 2 crosses node A in time interval 4.

Now look at the node crossing times of those two platoons at node D. Equation (3.6) determines the fraction of platoon 2 crossing node D in time interval 7. This fraction equals the elapsed time between the starting time of interval 7 and the node D crossing time of the end of platoon 2 (represented by dash line in Figure 3.1), divided by the elapsed time between the node D crossing times of platoons 1 and 2. Equation (3.6) is designed for $k = 0$. The need to take min[1, etc.] is that this calculation can exceed 1 when computing this fraction for platoons departing in time interval 1, since there is no node crossing time for a previous platoon.

Equation (3.7) determines the fraction of platoon 2 crossing node D in the *whole* time interval (if any) between the node D crossing times of platoons 1 and 2. The *whole* time interval is interpreted as follows. Suppose platoon 1 crosses node D in time interval 5 and platoon 2 crosses node D in interval 7 because delays have caused vehicles in platoon 2 to fall farther behind platoon 1. Under this situation, a fraction of platoon 2 crosses node D
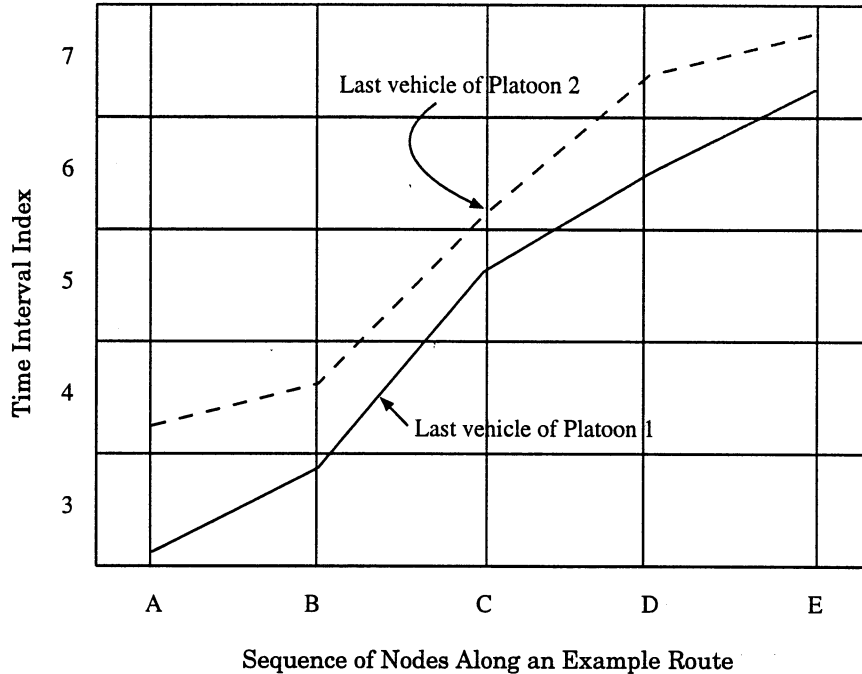
24

Figure 3.1: Effect of Flow Propagation Constraints

in interval 5, another fraction of platoon 2 crosses node D with *whole* time interval 6 and another fraction of platoon 2 crosses node D in interval 7. In this case, however, there are no so-called *whole* time intervals.

Equation (3.8) determines the fraction of platoon 2 crossing node D in time interval 6. This fraction equals the elapsed time between the crossing time of node D of the last vehicle in platoon 1 (represented by solid line) and the starting time of interval 7, divided by the elapsed time between the crossing times of node D of platoons 1 and 2. Equation (3.6) is designed for $k = 1$. The need to take max[0, etc.] in Equation (3.8) is that this fraction can be negative when computing this fraction for platoons departing in time interval 1, since there is no node crossing time for a previous platoon.

Note that $\phi_{ri}^d[t]$, $\alpha_{ri}^d[t]$ and $\pi_{ri}^d[t]$ used in Equations (3.6)–(3.8) are not solved simultaneously. They are updated iteratively once $\alpha_{ri}^d[t]$ and $\pi_{ri}^d[t]$ are initialized by the starting flow pattern, and $\phi_{ri}^d[t]$ is revised according to Equations (3.6)–(3.8). See Section 4.2 for additional details and Section 4.3 for the procedure of updating $\alpha_{ri}^d[t]$ and $\pi_{ri}^d[t]$.

An issue to be considered is whether shorter time intervals can be adopted to approximate continuous flow propagation instead of using the *flow fractions* defined by Equations (3.6)–(3.8). The answer is a tentative yes. However, shorter time intervals definitely require more computational effort to solve the model. First, more time intervals spanning the same analysis period require more calculations than using a longer time interval, plus the additional constraints of *flow fractions*. Second, FIFO ordering of trips in this discrete

time formulation is best maintained if link travel times remain well below the length of time interval. Therefore, using shorter length of time intervals may involve dividing links into shorter links which increases the effort of network coding and computation. Using *flow fractions* defined by Equations (3.6)–(3.8) converts a discrete time formulation into a quasi-continuous model.

3.2.4 First-In-First-Out Constraints

Although FIFO conditions may or may not occur in the real world, FIFO conditions should be strictly maintained in a dynamic route choice model. If FIFO conditions are violated for any link, a late entering trip flow will propagate faster than an earlier entering flow. Thus, temporally-correct flow propagation is not maintained for all O-D pairs. For a continuous time model, the flow propagation constraints imply FIFO constraints if a rigorous travel time function is used in the flow propagation constraints to ensure flow $x_a[t]$ on link a with travel time $\tau_a(x_a[t])$ regardless of the O-D source of this inflow. For a discrete or a quasi-continuous time model, however, FIFO conditions need to be explicitly stated to prevent potential violations.

Equations (3.9)–(3.12) state FIFO constraints between all O-D pairs according to their travel times in successive time intervals. Define β_{ri}^d as the time at which the last flow departing zone r in time interval d crosses node i via its shortest route less FIFO delay time at node i; $\theta_{ri}^d[t]$ is defined as the fraction of a time interval t that the last flow departing zone r in time interval d crosses node i; and $\mu_{ra}^d[t]$ is defined as the average travel time on link a of the last flow departing zone r in time interval d. The value h ($0 \leq h \leq 1$) is the fraction of a time interval that the end of the platoon (the last vehicle) departing from zone r in time interval d must follow the end of the platoon departing from zone r in time interval $d-1$. Vehicles are assumed to make one-for-one (or zero-sum) exchanges of traffic positions along any link, which is an acceptable and expected feature for any aggregate traffic model (Janson and Robles, 1995).

$$\pi_{ri}^d = \max\left(\beta_{ri}^d \,,\; \pi_{ri}^{d-1} + h\Delta t\right) \quad \forall\, r,i,d \;\; \text{and} \;\; \pi_{ri}^0 = \pi_{ri}^1 - \Delta t \tag{3.9}$$

$$\theta_{ri}^d[t] = \left[\left(\pi_{ri}^d - (t-1)\Delta t\right)/\Delta t\right]\alpha_{ri}^d[t] \quad \forall\, r,i,d,t \tag{3.10}$$

$$\mu_{ra}^d[t] = \left[\theta_{ri}^d[t]\tau_a(x_a[t]) + (1-\theta_{ri}^d[t])\tau_a(x_a[s])\right]\alpha_{ri}^d[t] \quad \forall\, r,a,d,t, \;\; s = t-1 \tag{3.11}$$

$$\left\{\beta_{rj}^d - \max\left[\pi_{ri}^d \,,\; (t-1)\Delta t + \Delta\tau_a^t[s]\right]\right\}\alpha_{ri}^d[t] \leq \mu_{ra}^d[t]\alpha_{ri}^d[t] \tag{3.12}$$

$$\forall\, r,a,d,t, \;\; s = t-1; \quad \text{where} \;\; \Delta\tau_a^t[s] = \tau_a(x_a[s]) - \mu_{ra}^d[t]$$

Equation (3.9) is a vehicle following constraint that regulates flows departing from the same zone in successive time intervals from passing each other. When solving for π_{ri}^d on the left hand side of Equation (3.9), π_{ri}^{d-1} on the right hand side is held fixed. If $h = 0$, a trailing

platoon can completely overlay (but not overtake) a leading platoon so that the two platoons become coincident, which is not realistic. If $h = 1$, a trailing platoon can never partly gain ground on a leading platoon.

Since Equation (3.9) does not insure FIFO ordering between all O-D pairs, Equations (3.10)–(3.12) are required. Equations (3.10) and (3.11) determine the average travel time on link a of the end of the platoon departing zone r in time interval d adjusted for the time into interval t versus $t - 1$ that the platoon enters the link. Equations (3.10) and (3.11) dampen speed transitions between time intervals in a quasi-continuous manner so that vehicle speeds do not abruptly change if flows enter links just seconds before or after a time interval change.

In most cases, the link travel time is well below the duration of each time interval (Δt). However, Equation (3.12) is needed to prevent FIFO violations in cases where link travel times exceed Δt. Equation (3.12) does not entirely replace the need for Equation (3.9). Equation (3.12) allows trips between different O-D pairs to become concurrent while sharing the same route. Equation (3.9) insures a minimum separation of the last platoon departing from the same zone in successive time intervals. Trips from the same zone bunch together and cause excessively dense flows if Equation (3.9) is removed. By using Equations (3.9)–(3.12), the FIFO conditions are maintained in a quasi-continuous manner for the proposed VI route choice model.

3.2.5 Nonnegativity Constraints

Finally, all variables must be nonnegative at all time intervals. We have

$$x_a^{rs}[t] \geq 0 \qquad \forall\, r, s, a, t; \tag{3.13}$$

$$e_{rs}[t] \geq 0 \qquad \forall\, r, s, t; \tag{3.14}$$

$$\phi_{ri}^d[t] \geq 0 \qquad \forall\, r, i, d, t. \tag{3.15}$$

3.3 Link-Time-Based Conditions

The link-time-based VI model of the DUO route choice problem is proposed to solve for the travel-time-based *ideal* DUO state defined in Section 3.1.2. In the case of an *ideal* DUO state, the equilibration of route travel times is stated for each O-D pair instead of each decision-node/destination pair because the *ideal* DUO focuses on the optimal state along the entire journey.

We now derive the equivalent mathematical inequalities for the travel-time-based *ideal* DUO state using link variables. For any route from origin r to destination s, link a is defined as being used in time interval t if $x_a^{rs}[t] > 0$. Define π_{ri}^{d*} as the minimal travel time actually experienced by flows departing from origin r to node i in time interval d, the asterisk

denoting that the travel time is calculated using DUO traffic flows. For link $a = (i,j)$, the minimal travel time π_{rj}^{d*} from origin r to node j should equal to or less than the minimal travel time π_{ri}^{d*} from origin r to node i plus the actual link travel time $\tau_a[d + \bar{\pi}_{ri}^d]$ in time interval $[d + \bar{\pi}_{ri}^d]$, where the $\bar{\pi}_{ri}^d$ denotes the number of time intervals traversed in π_{ri}^d; see definition in Section 3.2.2 and Equation (3.2). The first time interval of $[d + \bar{\pi}_{ri}^d]$ must be the earliest time interval that flow departing zone r in time interval d can enter link a. It follows that

$$\pi_{ri}^{d*} + \tau_a[d + \bar{\pi}_{ri}^{d*}] \geq \pi_{rj}^{d*} \qquad \forall\, a = (i,j), r, d. \tag{3.16}$$

If, for each O-D pair (r,s), any departure flow from origin r in time interval d enters link a at the earliest time interval $[d + \bar{\pi}_{ri}^{d*}]$, or $x_a[d + \bar{\pi}_{ri}^{d*}] \geq 0$, then the *ideal* DUO route choice conditions require that link a is on the route with minimal travel time. In other words, the minimal travel time π_{rj}^{d*} from origin r to node j should equal to the minimal travel time π_{ri}^{d*} from origin r to node i plus the actual link travel time $\tau_a[d + \bar{\pi}_{ri}^{d*}]$ in time interval $[d + \bar{\pi}_{ri}^{d*}]$. It follows that

$$\pi_{rj}^{d*} = \pi_{ri}^{d*} + \tau_a[d + \bar{\pi}_{ri}^{d*}], \quad \text{if } x_a^{rs*}[d + \bar{\pi}_{ri}^{d*}] > 0 \quad \forall\, a = (i,j), r, s, d. \tag{3.17}$$

The above equation is also equivalent to the following:

$$x_a^{rs*}[d + \bar{\pi}_{ri}^{d*}] \left[\pi_{ri}^{d*} + \tau_a[d + \bar{\pi}_{ri}^{d*}] - \pi_{rj}^{d*} \right] = 0 \quad \forall\, a = (i,j), r, s, d. \tag{3.18}$$

Thus, the link-time-based *ideal* DUO route choice conditions can be summarized as below:

$$\pi_{ri}^{d*} + \tau_a[d + \bar{\pi}_{ri}^{d*}] - \pi_{rj}^{d*} \geq 0 \quad \forall\, a = (i,j), r; \tag{3.19}$$

$$x_a^{rs*}[d + \bar{\pi}_{ri}^{d*}] \left[\pi_{ri}^{d*} + \tau_a[d + \bar{\pi}_{ri}^{d*}] - \pi_{rj}^{d*} \right] = 0 \quad \forall\, a = (i,j), r, s, d; \tag{3.20}$$

$$x_a^{rs}[d + \bar{\pi}_{ri}^{d*}] \geq 0 \quad \forall\, a = (i,j), r, s, d. \tag{3.21}$$

3.4 The Link-Time-Based VI Model

Define Ω_{ra}^{d*} as the difference between the minimal travel time from zone r to node i plus the travel time on link a and the minimal travel time from zone r to node j for flows departing from zone r in time interval d.

$$\Omega_{ra}^{d*} = \pi_{ri}^{d*} + \tau_a[d + \bar{\pi}_{ri}^{d*}] - \pi_{rj}^{d*} \tag{3.22}$$

The link-time-based *ideal* DUO route choice conditions are rewritten as:

$$\Omega_{ra}^{d*} \geq 0 \qquad \forall\, a = (i,j), r, d; \tag{3.23}$$

$$x_a^{rs*}\left[d + \bar{\pi}_{ri}^{d*}\right] \Omega_{ra}^{d*} = 0 \qquad \forall\, a = (i,j), r, s, d; \tag{3.24}$$

$$x_a^{rs}\left[d + \bar{\pi}_{ri}^{d*}\right] \geq 0 \qquad \forall\, a = (i,j), r, s, d. \tag{3.25}$$

28

The equivalent VI formulation of the link-time-based *ideal* DUO route choice conditions defined in Equations (3.23)–(3.25) can now be stated as follows.

$$\sum_d \sum_{rs} \sum_a \Omega_{ra}^{d*} \left\{ x_a^{rs}[d + \bar{\pi}_{ri}^{d*}] - x_a^{rs*}[d + \bar{\pi}_{ri}^{d*}] \right\} \geq 0 \tag{3.26}$$

where $*$ denotes the DUO state, and where the dynamic traffic flow pattern must satisfy the constraints described in Equations (3.1)–(3.15).

3.4.1 Proof of Necessity

We need to prove that link-time-based *ideal* DUO route choice conditions defined in Equations (3.23)–(3.25) imply to variational inequality (3.26). For link a, a feasible inflow in time $[d + \bar{\pi}_{ri}^{d*}]$ is

$$x_a^{rs}[d + \bar{\pi}_{ri}^{d*}] \geq 0 \tag{3.27}$$

Multiplying Equation (3.27) and the DUO route choice condition (3.23), we have

$$x_a^{rs}[d + \bar{\pi}_{ri}^{d*}] \Omega_{ra}^{d*} \geq 0 \qquad \forall\, a = (i,j), r, s, d. \tag{3.28}$$

Next, subtract the *ideal* DUO route choice condition (3.24) from Equation (3.28) and obtain

$$\left\{ x_a^{rs}[d + \bar{\pi}_{ri}^{d*}] - x_a^{rs*}[d + \bar{\pi}_{ri}^{d*}] \right\} \Omega_{ra}^{d*} \geq 0 \qquad \forall\, a = (i,j), r, s, d. \tag{3.29}$$

Summing Equation (3.29) for all links a ($a \in \mathcal{A}$) and O-D pairs (r,s) $(r, s \in \mathcal{Z})$, it follows that

$$\sum_{rs} \sum_a \left\{ x_a^{rs}[d + \bar{\pi}_{ri}^{d*}] - x_a^{rs*}[d + \bar{\pi}_{ri}^{d*}] \right\} \Omega_{ra}^{d*} \geq 0 \tag{3.30}$$

Summing Equation (3.30) for all time intervals, the following variational inequality is obtained and is identical to Equation (3.26).

$$\sum_d \sum_{rs} \sum_a \Omega_{ra}^{d*} \left\{ x_a^{rs}[d + \bar{\pi}_{ri}^{d*}] - x_a^{rs*}[d + \bar{\pi}_{ri}^{d*}] \right\} \geq 0 \tag{3.31}$$

Q.E.D.

3.4.2 Proof of Sufficiency

For any solution $x_a^{rs*}[d + \bar{\pi}_{ri}^{d*}]$ to variational inequality (3.26), we need to prove that it satisfies the link-time-based *ideal* DUO route choice conditions defined in Equations (3.23)–(3.25). Since Equations (3.23) and (3.25) must hold by definition, only the equivalence of Equation (3.24) needs to be proved.

Assume that the second *ideal* DUO route choice condition (3.24) does not hold only for a link $\hat{a} = (\hat{i}, \hat{j})$ for O-D pair $(\hat{r}, \hat{s})$ during time interval $[d_1 - \delta, d_1 + \delta] \in [0, \mathcal{T}]$ where δ denotes the specific time interval. Thus,

$$x_{\hat{a}}^{\hat{r}\hat{s}*}\left[d + \bar{\pi}_{\hat{r}\hat{i}}^{d*}\right] > 0 \quad \text{and} \quad \Omega_{\hat{r}\hat{a}}^{d*} > 0 \tag{3.32}$$

In other words, we have

$$x_{\hat{a}}^{\hat{r}\hat{s}*}\left[d + \bar{\pi}_{\hat{r}\hat{i}}^{d*}\right]\Omega_{\hat{r}\hat{a}}^{d*} > 0 \tag{3.33}$$

Since the second *ideal* DUO route choice condition (3.24) holds for all cases other than link $\hat{a} = (\hat{i}, \hat{j})$ for O-D pair $(\hat{r}, \hat{s})$ during time interval $[d_1 - \delta, d_1 + \delta]$, it follows that

$$\sum_d \sum_{rs} \sum_a \Omega_{ra}^{d*}\, x_a^{rs*}\left[d + \bar{\pi}_{ri}^{d*}\right] = \sum_d \Omega_{\hat{r}\hat{a}}^{d*}\, x_{\hat{a}}^{\hat{r}\hat{s}*}\left[d + \bar{\pi}_{\hat{r}\hat{i}}^{d*}\right] > 0 \tag{3.34}$$

Note all other terms in Equation (3.34) have vanished because of *ideal* DUO route choice condition (3.24).

For each O-D pair (r, s), a minimal travel time route p for flow departing origin r in time interval d can always be found, where route p is evaluated under the DUO flow pattern, $x_a^{rs*}\left[d + \bar{\pi}_{ri}^{d*}\right]$. For this route p, the first *ideal* DUO route choice condition (3.23) becomes an equality by definition. That is

$$\Omega_{ra}^{d*} = 0 \qquad \forall\, a = (i,j), r;\; a \in p. \tag{3.35}$$

Next, we need to find a set of feasible inflows $x_a^{rs}\left[d + \bar{\pi}_{ri}^{d*}\right]$ so that Equation (3.36) holds.

$$x_a^{rs}\left[d + \bar{\pi}_{ri}^{d*}\right]\Omega_{ra}^{d*} = 0 \qquad \forall\, a = (i,j), r, s, d. \tag{3.36}$$

For each O-D pair (r, s) at each time interval d, we assign O-D departure flow $f_{rs}[d]$ to the minimal travel time route p, which was evaluated under the DUO flow pattern, $x_a^{rs*}\left[d + \bar{\pi}_{ri}^{d*}\right]$. A set of feasible inflow pattern $x_a^{rs}\left[d + \bar{\pi}_{ri}^{d*}\right]$ that always satisfies Equations (3.35) and (3.36) is generated because flows are never assigned to routes without minimal travel times. Summing Equation (3.36) for all links a and all O-D pairs (r, s), we obtain

$$\sum_{rs} \sum_a x_a^{rs}\left[d + \bar{\pi}_{ri}^{d*}\right]\Omega_{ra}^{d*} = 0 \qquad \forall\, a = (i,j), r, s, d. \tag{3.37}$$

Summing Equation (3.37) for all time intervals, we have

$$\sum_d \sum_{rs} \sum_a x_a^{rs}\left[d + \bar{\pi}_{ri}^{d*}\right]\Omega_{ra}^{d*} = 0 \tag{3.38}$$

We subtract Equation (3.34) from Equation (3.38) and get

$$\sum_d \sum_{rs} \sum_a \Omega_{ra}^{d*}\left\{x_a^{rs}[d + \bar{\pi}_{ri}^{d*}] - x_a^{rs*}[d + \bar{\pi}_{ri}^{d*}]\right\} \leq 0 \tag{3.39}$$

Q.E.D.

The above equation contradicts variational inequality (3.26). Therefore, any optimal solution $x_a^{rs*}[d + \bar{\pi}_{ri}^{d*}]$ of Equation (3.26) satisfies the second *ideal* DUO route choice condition (3.24).

3.5 A Combination of the Instantaneous and Ideal DUO States

In the *ideal* DUO route choice problem, travelers are assumed to have perfect information about the future network conditions at the times of their departures. This strong assumption benefits DUO route choice modeling because route choice decisions must be based on travel times which are temporally-consistent with future link flows at time of link use for an analytical-based route choice model. On the contrary, this assumption complicates the modeling of many realistic route choice options. The assumption is appropriate for recurrent trips and traffic conditions, scheduled events (e.g., detour, road construction) and even for predicted weather conditions. However, this behavioral assumption is inconsistent with *unexpected* capacity reducing events (e.g., accidents, stalled vehicles) at future times because drivers have very limited capability to be informed about the times and locations of such events before they encounter unusual queuing delays caused by those incidents. Enroute diversions are thus expected to happen when incidents are encountered by drivers.

As for the *instantaneous* DUO route choice problem, travelers always choose routes according to currently prevailing traffic conditions. This behavioral assumption may be unrealistic because travelers may not change routes continuously but can do so when unusual congestion caused by unexpected capacity reducing events are encountered. Therefore, travelers may choose to divert to avoid congestion if *instantaneous* route choice is applied.

In this thesis, the proposed model described in Sections 3.2 and 3.3 is used to solve the *ideal* DUO route choice problem without considering any capacity reducing events. However, alternative route choice strategies based on both anticipatory (*ideal* DUO) and non-anticipatory (*instantaneous* DUO) traffic conditions can be combined to limit the possible incident impacts between the least and most severe cases in the algorithm when solving the model. Detailed modeling issues and approaches related to this issue are presented in Section 5.6.

Chapter 4

Solution Algorithm

The algorithm used for solving the proposed link-time-based VI model of DUO route choice is stated in this chapter. Static route choice problems can be solved efficiently by convex combinations methods (e.g., the Frank-Wolfe (F-W) algorithm) for nonlinear programs with linear constraints. These methods, however, may create temporally discontinuous flows when directly applied to dynamic route choice problems. Instead, a diagonalization (or double relaxation) algorithm is proposed to solve the VI DUO route choice model to convergence.

This algorithm consists a sequence of *diagonalization* iterations. First, in the initialization step, the *node time intervals* and *shortest route travel times* are initialized based on the initial link flows. Then, the algorithm solves a sequence of route choice problems (called *inner* problems) using the F-W algorithm with fixed *node time intervals* until the convergence criterion of the route choice problem is satisfied. Next, the *node time intervals* and *shortest route travel times* are updated (called *outer* problems) based on the most recently assigned link flows from the *inner* problem. Adjustments of link capacities are made between the *inner* and *outer* problems to account for capacity changes caused exogenously (e.g., signal timing changes, incidents and other unexpected capacity reduction events) or generated endogenously (e.g., queue spillbacks). The algorithm terminates when the changes of the *node time intervals* obtained from two consecutive *outer* problems are within a prespecified tolerance. Figure 4.1 shows the steps of the solution algorithm, which are described in detail in the following sections.

4.1   Initialization

In the initialization step, we need to do the following:

1. input the O-D matrix, trip departure rates, network data, and initial link flows;

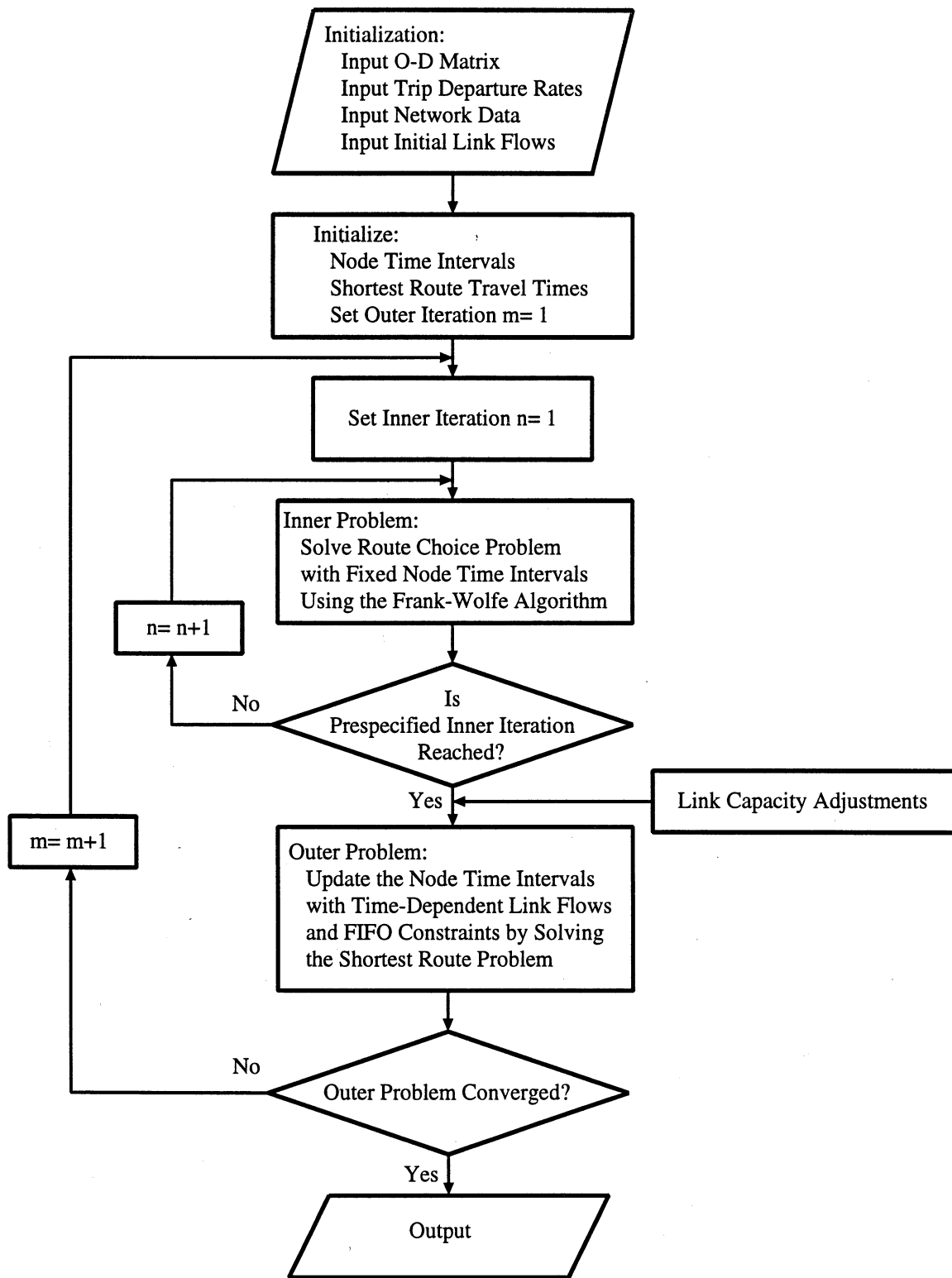2. initialize the *shortest route travel times* and *node time intervals.*

Figure 4.1: Flowchart of the Solution Algorithm

Depending on data availability, the O-D matrix usually covers a much longer time period (e.g., 1–hour or 24–hour) than the length of time interval (typically from 5 minutes to 15 minutes) used in the solution procedure. Consequently, departure rates for each zone and each time interval to convert the O-D matrix from a longer time period into shorter time intervals are necessary. As for the network data, link-node incidences, free-flow travel times, link capacities and other data that fulfill the need for calculating link travel times such as green-splits and signal cycle lengths are needed. Initial link flows are optional, and can be set to zero, but the static user-optimal flows scaled down to the duration of the selected time interval might be good starting values. Based on the initial link flows $(x_a[t])$, the *shortest route travel times* (π_{ri}^d) and *node time intervals* $(\alpha_{ri}^d[t])$ are initialized by solving a shortest route problem (see Section 4.3 for details).

4.2 Solving the Route Choice Problem

According to the optimal values of *shortest route travel times* (π_{ri}^d) and *node time intervals* $(\alpha_{ri}^d[t])$, *flow fractions* $(\phi_{ri}^d[t])$ departing zone r in departure time interval d that cross node i in time interval t are computed using Equations (3.6)–(3.8). Once the *flow fractions* are updated using the *node time intervals* and *shortest route travel times*, they are held fixed when the objective function is minimized subject to Equations (3.1)–(3.8) and Equations (3.13)–(3.15). The *flow fractions* are required to maintain temporally-correct routes and time-continuous flow propagations over time intervals. This step can be viewed as solving a sequence of static route choice problems. The algorithm of Frank and Wolfe (1956), or other appropriate convex combinations method, can be applied to solve this step to convergence. The steps of the F-W algorithm to solve the *inner* problem are described as follows.

Step 0: Initialization. Perform an all-or-nothing assignment based on fixed *node time intervals, flow fractions* and initial minimal-cost (travel time) route for each O-D pair in each departure time interval. This yields a main-problem solution $(x_a^n[t])$ of the *inner* problem. Set *inner* iteration $n = 1$.

Step 1: Update. Recalculate the link travel times $\tau_a^n(x_a[t])$ for each link in each time interval based on the assigned link flows.

Step 2: Direction Finding. Based on the updated link travel times $\tau_a^n(x_a[t])$, search minimal-cost (travel time) routes for each O-D pair in each departure time interval. Perform an all-or-nothing assignment, yielding a subproblem solution $(y_a^n[t])$ of the *inner* problem.

Step 3: Line Search. Find an optimal step size λ $(0 \leq \lambda \leq 1)$ that solves the one-dimensional search problem.

$$\min_{0 \leq \lambda \leq 1} \sum_t \sum_a \int_0^{(1-\lambda)x_a^n[t]+\lambda y_a^n[t]} \tau_a(x_a[t]) \, dx \qquad (4.1)$$

The bisection method is selected to find the optimal step size λ of the *inner* problems. The integration of the objective function is avoided if the bisection method is adopted.

Step 4: Move. Using the optimal step size, find a new solution by combining the main-problem and subproblem solution of the *inner* problem.

$$\text{Set} \quad x_a^{n+1}[t] = (1 - \lambda)x_a^n[t] + \lambda y_a^n[t] \quad \forall a, t. \qquad (4.2)$$

Step 5: Convergence Test for Inner Iterations. If n equals a prespecified iteration number, go to the *outer* problem (Section 4.3); otherwise, set $n = n + 1$ and go to Step 1 of the *inner* problem.

4.3   Updating the Node Time Intervals

The purpose of updating the *node time intervals* is to maintain the temporally-correct routes and time-continuous traffic flow propagations in successive time intervals as well. In each *outer* iteration, the *node time intervals* are updated by solving the shortest route problem with time-dependent link travel times obtained from the *inner* problem (route choice problem). The adopted shortest route algorithm has been modified for dynamic problems so that the link travel times used in pivoting from a node depend on the time interval in which the shortest route tree departs from that node and the link flows in that time interval.

To update the *node time intervals* $(\alpha_{ri}^d[t])$, the following procedure is applied:

1. find the shortest route travel times (π_{ri}^d) and time intervals $(\bar{\pi}_{ri}^d)$;

2. reset values of $(\alpha_{ri}^d[t])$ as follows:

> if $(\bar{\pi}_{ri}^d \leq t\Delta t)$ and $(\bar{\pi}_{ri}^d \geq (t-1)\Delta t)$, then $\alpha_{ri}^d[t] = 1$;
>
> otherwise $\alpha_{ri}^d[t] = 0$.
>
> perform for all $r \in \mathcal{Z}$, $i \in \mathcal{N}$; $d = 1, \cdots T$ and $t = d, d+1, \cdots T$
>
> note: $\sum_{t=d}^{T} \alpha_{ri}^d[t] = 1$;

3. set $\pi_{rr}^d = d\Delta t, \; \forall r \in \mathcal{Z}, \; d \in \mathcal{T}$;

4. enforce the first-in-first-out (FIFO) conditions stated in Equations (3.9)–(3.12).

Note π_{rr}^d equals the start time of the end of the platoon departing zone r in time interval d. π_{rr}^d is set to $d\Delta t$ in order to set the clock correctly to the end of each time interval. Although $\alpha_{ri}^d[t]$ can never equal 1 when $t = 1$, flows departing in interval 1 are uniformly distributed over the previous time span $(\bar{\pi}_{ri}^1 - \Delta t)$ at each node of the network such that some flows are still assigned in time interval 1.

4.4 Adjustment of Link Capacities

A key feature of this model is that link-specific capacity adjustments are explicitly considered in the solution process to improve further the reality of the solution. Adjustments of link capacities are made between the *inner* and *outer* problems to account for capacity changes caused exogenously or generated endogenously.

Exogenous link capacity adjustments can be specified to the program when they are detected by the roadway surveillance systems (e.g., detectors, CCTV), reported by the drivers and/or highway patrols and scheduled to occur. The capacity adjustment events include traffic accidents, stalled vehicles, dropped objects, dangerous chemical spills, weather effects, changes of signal timing plans, roadway constructions, time-of-day roadway usage restrictions (e.g., reversible lanes), etc. Endogenous link capacity adjustments are usually caused by spillback queues that reduce capacities of upstream links. However, if mechanisms of ramp metering and/or advanced traffic signal systems (e.g., actuated, adaptive and responsive signal timing schemes) are presented or integrated into the model, the caused endogenous capacity adjustments can be captured as well.

Adjustments of link capacities are made on inflow links to nodes whose outflow links have flow greater than γ times of the original capacity ($\gamma = 1.05$ is being used). If queues on outflow links have spilled back to their tail node (at the intersection), then speeds on inflow links should approximate the weighted speed of outflow links. The fraction of link a that is blocked by a queue within time interval t, (ψ_a^t), must also be measured. If ψ_a^t is smaller than 1, then there is no effect of inflows on link a in time interval t. If ψ_a^t is greater than 1, then the portion of the time interval of the inflow link affected by this queue is determined. In addition, whether the queue is decreasing or increasing within the time interval is also decided. The essential steps of these capacity adjustments and queue propagations are:

1. track the location of multiple queue spillbacks in the network;

2. weight the effects of multiple queue spillbacks that jointly affect inflows to any node;

3. adjust the capacities of the inflow links to each node in proportion to the fractions of flows affected by each queue.

Based on Janson and Robles (1995), the procedure of queue propagation and capacity adjustment is described next. Note that all nodes in the network representation are configured such that:

1. For each *merge* node, there is only one outflow link.

2. For each *diverge* node, there is only one inflow link.

Therefore, no node has multiple inflow and outflow links simultaneously (i.e., no node is both a merge node and a diverge node). Intersections always have turning movement links, and weaving sections on freeways always have connecting links between the diverge and merge nodes. Steps of capacity adjustment starting from original unadjusted capacities are:

Step 1. Increment the iteration number from 1 until changes of link capacities of all links in all time intervals are stabilized.

Step 2. Increment the time interval index from 1 to $\mathcal{T}$.

Step 3. Increment the node number from 1 to $\mathcal{N}$.

Step 4. For each outflow link:

1. Compute the cumulative queue equal to all excess flow exceeding the capacity through the current time interval.

2. The fraction (ψ_a^t) of link length occupied by queue during the current time interval (say t) is computed as the cumulative queue divided by the absorbable flow of link a. If $\psi_a^t < 1$, then there is no effect on inflows to link a in time interval t. If $\psi_a^t > 1$, then the fraction of time interval of the affected inflows depends on when this condition occurred, and whether this queue is increasing or decreasing. The absorbable flow of link a is computed as:

    absorbable flow of link a = (density^2 - density^1) × (length of link a)

    where density^1 = flow^1/speed^1 = density before the queue (less dense).

    density^2 = flow^2/speed^2 = density inside the queue (more dense).

    flow^1 = flow before the queue (below original but above adjusted capacity).

    flow^2 = flow inside the queue (assumed to equal adjusted capacity).

    speed^1 = link length divided by travel time for flow^1 with original capacity.

    speed^2 = link length divided by travel time for flow^2 with adjusted capacity.

37

Step 5. Inflow links are unaffected until a queue extends beyond the tail node of an outflow link (i.e., $\psi_a^t > 1$), and only a fraction of time interval will be affected when this queue spillback first occurs. The affected time interval fraction equals (the interval starting time)–(time that the queue reaches the tail node of outflow link) divided by the duration of each time interval (Δt).

Step 6. Compute the weighted flow-to-capacity ratio for each affected outflow link. For each affected outflow link, this ratio is weighted by the flow and the affected time interval fraction found in Step 5.

Step 7. Adjust the capacity of each inflow link so that the corresponding flow-to-capacity ratio equals the weighted flow-to-capacity ratio of the outflow links found in Step 6. Return to Step 3 until all nodes processed, then return to Step 2 until all time intervals processed. If the capacity of any link in any time interval changes more than the prespecified percent, return to Step 1 for the next iteration; otherwise, stop the procedure of link capacity adjustment.

In order to capture the effects of potential multiple queues from different locations in the network and queues spilling back farther than one link in any time interval, multiple iterations of the capacity adjustment procedure are performed until the capacity of each link in all time intervals do not change significantly. Constraints are added to prevent endogenously adjusted capacity from becoming too small and to prevent queues from dissipating too quickly. That is, during the queue dissipation, the link capacity cannot exceed the adjusted capacity and its reduced capacity in the prior time interval. Then, adjusted capacities are returned to the *inner* problem stated in Section 4.2. Consequently, both link flows and capacities are time-dependent in the adopted link travel time functions.

4.5 Convergence Test of the Outer Iteration

The last step of the solution algorithm is the convergence test. The convergence index of the solution algorithm is defined as the changes of the *node time intervals* between two consecutive *outer* iterations of the algorithm. The solution algorithm terminates if the changes of the *node time intervals* (NDIFFS) are less than *nodes* $\times$ *zones* $\times$ *intervals* $\times$ *x*. This index indicates that only x changes of the total node time intervals are allowed for the last flow in the platoon departing from each zone over the total analysis period. A sensitivity analysis of the x-value might be needed. With the perfect convergence, the changes of the *node time intervals* equal zero.

4.6　Summary of the Solution Algorithm

The steps of the solution algorithm of the proposed link-time-based variational inequality dynamic user-optimal route choice model are summarized as follows.

Step 1: Initialization. Input the O-D matrix, time-dependent trip departure rates, network data and initial link flows (optional). With the initial link flows, the *shortest route travel times* and *node time intervals* are initialized as well. Set *outer* iteration $m = 1$.

Step 2: Solve the Route Choice Problem (Inner Problem). Using the F-W algorithm, the route choice problem is solved to convergence with the optimal values of the *node time intervals* from the *outer* problem.

Step 3: Link Capacity Adjustments. Perform exogenous and endogenous link capacity adjustments.

Step 4: Update the Node Time Intervals (Outer Problem). Update the *node time intervals* with time-dependent link flows and FIFO constraints by solving the shortest route problem.

Step 5: Convergence Test for Outer Iterations. Sum the total number of differences of the *node time intervals* (NDIFFS) between iterations $m - 1$ and m. If NDIFFS $\leq$ allowable percentage of all *node time intervals*, then the algorithm stops; otherwise, return to Step 2.

Chapter 5

Model Implementation

Issues of model implementation are addressed in this chapter. The test network, network representation and adopted link travel time functions are described. The contents extend to the preparation of required data including travel demand, time-dependent trip departure rates and roadway facility attributes such as types of links, types of turning movements and types of intersection control, etc. A notable extension is that this model is capable of modeling enroute diversions resulting from incidents. Issues and modeling approaches to enroute diversion are presented as well.

5.1   Test Network

The test network selected for testing the proposed dynamic user-optimal (DUO) route choice model is the ADVANCE Network. ADVANCE (Advanced Driver and Vehicle Advisory Navigation Concept), a field test of ATIS was recently concluded by the Illinois Department of Transportation (IDOT) and the Federal Highway Administration (FHWA), in collaboration with the University of Illinois at Chicago, Northwestern University, and the IVHS Strategic Business Unit of Motorola, Inc. The ADVANCE Network is depicted in Figure 5.1. It is located in the northwestern suburbs of Chicago and covers about 300 square miles (800 square kilometers). Diversified land use patterns including dense residential communities, office centers, regional shopping centers, subregional government centers and the O'Hare International Airport are located in the ADVANCE Test Area. The topology of the test network is almost a regular grid with a few diagonal major arterials directed towards the Chicago CBD (Central Business District). The freeway system includes I-90, I-94, I-190, I-290, I-294, IL-20 and IL-53. Except for the remote northwest corner, the freeways serve nearly all parts of the test network. The southwest quadrant is characterized by modern, multi-lane arterials designed for high volumes. In Figure 5.1, collectors, arterials and freeways are drawn with lines of different widths, freeways being the widest line. The heavy black line indicates the
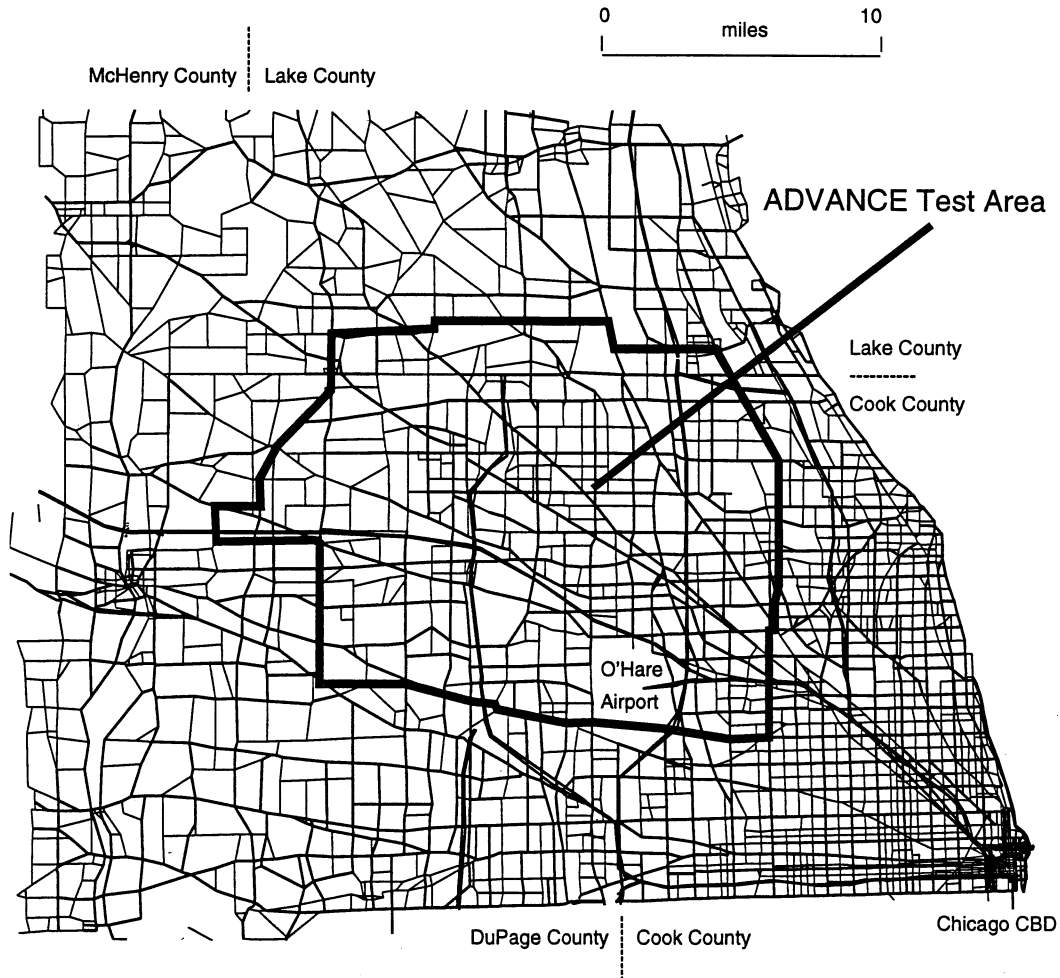
boundary of the ADVANCE Network.

Figure 5.1: The ADVANCE Test Area in the Northwestern Suburbs of Chicago

Table 5.1 lists the frequency of links by the facility type and related data. Table 5.2 presents the breakdown of arterial/collector intersections by the number of legs and types of traffic control.

5.2 Expanded Intersection Representation

In a conventional route choice model, the network is coded with each intersection represented as a single node and each approach is represented as a single link. This network representation defines travel times in terms of approach attributes and traffic flows. Using

41

Table 5.1: Network Characteristics

Type of Facility	No. of Links
Arterial/Collector	4,061
Tollway/Freeway	197
Freeway Ramp	202
Toll Plaza	14
Freeway Weaving Section	11
Centroid Connector Links	2,491
Approach Link	874
Total	7,850
Number of Nodes	2,552
Number of Zones	447

Table 5.2: Intersection Frequency by Number of Legs and Control Type

	Signalized	Priority	All-way-stop	Total
Three-leg	257	174	51	482
Four-leg	558	52	60	670
Five-leg and more	7	0	0	7
Total	822	226	111	1,159

the conventional network representation, 7,850 approach links and 2,552 nodes are included in the ADVANCE Network (Table 5.1).

One of the goals of this research is to provide dynamic and specific link travel time estimates for the network including turning movements. Therefore, a more appropriate network representation known as an *expanded intersection representation* is required. The expanded intersection representation consists of defining a special network representation so each turning movement is represented by a separate link called an *intersection link*. Road segments between intersections are represented by links called *non-intersection links*. More precisely, an approach node is defined for each approach to an intersection, and the number of intersection links originating from this node equals the number of possible turning movements. Similarly, an exit node is defined for each exit from an intersection. Non-intersection links connect an exit node of one intersection with an approach node of another. For example, a typical four-leg intersection with two-way approaches without any turning restrictions (U-turn excluded), four approach nodes, four exit nodes and twelve intersection links are required in this expanded network representation. The expanded intersection representation is described by Figure 5.2.
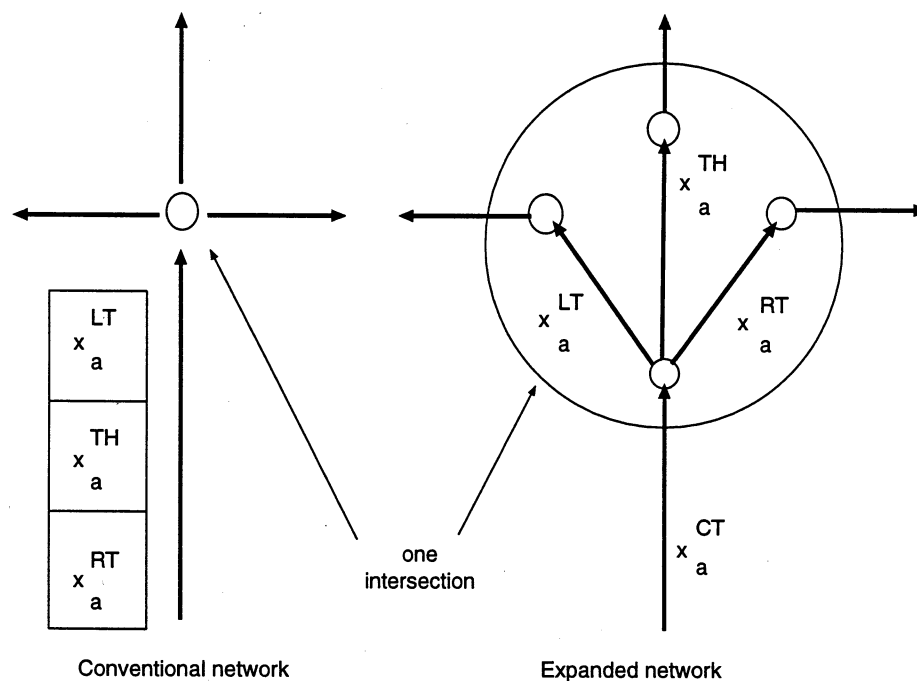
Figure 5.2: Expanded Intersection Representation

The expanded intersection representation procedure is applied only to nodes representing an intersection of arterials or collectors. Nodes that do not need to be expanded are freeway nodes, no-delay intersections, and other nodes not representing intersections. In the case of freeways, delays can be directly assigned to the link. No-delay intersections are those

intersections where no approach flow experiences any delay; for example, a diverge to a freeway ramp from an arterial. To keep this expanded intersection representation simple, all the turning restriction links have been removed from the network representation.

Because of the expanded intersection representation, the network size increases about three times in comparison with conventional network representation. To that end, 22,918 links and 9,700 nodes are actually modeled in solving the proposed DUO route choice model. Note the detailed delay functions by turning movements are applied only within the actual ADVANCE Network. The network expansions are performed by the network builder program that was originally developed by Meneguzzer et al. (1990) and enhanced by Berka et al. (1994).

5.3   Link Travel Time Functions

This section presents mathematical functions used within the route choice model to estimate link travel times for given flow rates. The choice of the delay functions involves several criteria:

1. the desired mathematical properties of the function to satisfy the condition for a unique solution of the model;

2. the cost and limited availability of road data;

3. the computational effort required by the model;

4. the desired accuracy of the travel time estimates generated by the model.

Considering the nature of this research, analytical functions are preferred over regression-based models because the former generate reasonable estimates over a much wider range of input flows and other parameters. Criteria (1) and (2) above exclude many highly detailed traffic engineering-based models. Criterion (3) excludes simulation models, which are more suitable for small networks.

Delay functions selected for this study can be classified by road type and intersection type. First, delay functions for signalized intersections are presented (Section 5.3.1); second for unsignalized intersections (Section 5.3.2); and third for freeway-related facilities (Section 5.3.3).

5.3.1   Signalized Intersections

In general, delay models for signalized intersections consist of three modules: (1) saturation flow analysis; (2) signal timing procedure; and (3) link travel time (delay) function. In this

44

research, signal timings and initial saturation flows are obtained from an asymmetric static user-optimal route choice model (Berka et al., 1994). During the solution procedure, the signal timings are held fixed and the saturation flows are adjusted if spillback queues and other capacity adjustment events (such as incidents) occur.

The specific delay function for links at signalized intersections applied in the research has the following form (Akcelik, 1988):

$$d = \frac{0.5C(1-u)^2}{1-ux} + 900T\gamma \left[x - 1 + \sqrt{(x-1)^2 + \frac{8(x-0.5)}{cT}} \right] \tag{5.1}$$

where d is the average delay per vehicle (second/vehicle), C is the signal cycle length (second), $u = g/C$ is the green split, g is the green time (second); $x = v/c$ is the flow-to-capacity ratio, T is the duration of the flow (hour) and $\gamma = 1$ for $x > 0.5$, and 0, otherwise. The first term, called the *uniform delay*, was originally proposed by Webster (1958). It reflects the average delay experienced by drivers in *undersaturation conditions*, that is when the arrival flow does not exceed capacity. In oversaturation conditions, $x = 1$ is used in the uniform delay term. The second term of Equation (5.1) is called the *overflow delay*. It reflects the delay experienced by the vehicles when the flow rate is close to or exceeds the capacity. Temporary overflow at an intersection may also occur when the average arrival rate is lower than the capacity, due to a random character of the arrival pattern. The earliest delay functions (for example, Webster, 1958) were based on the steady-state model and were defined only for undersaturation conditions. Figure 5.3 shows an example of an Akcelik function.
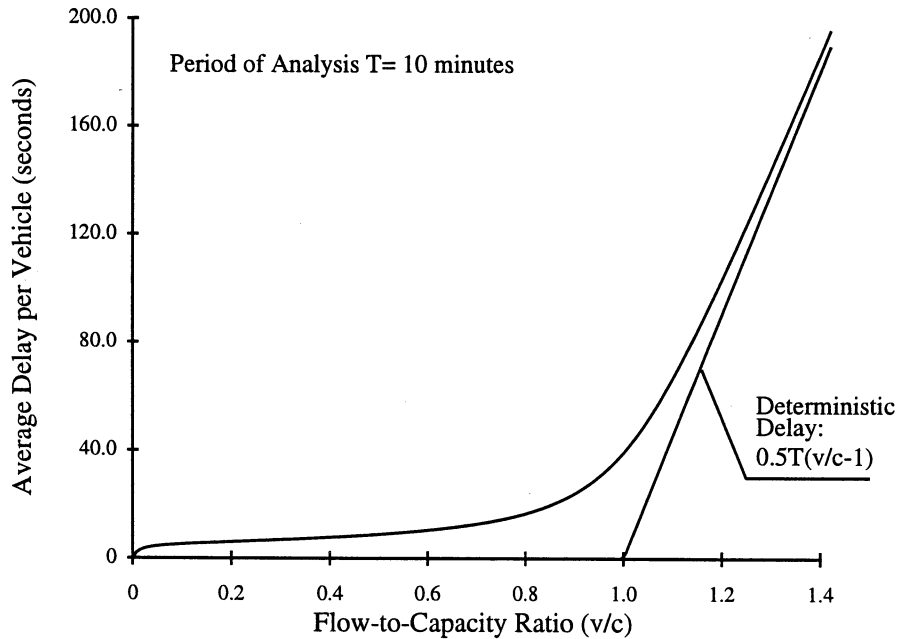
Figure 5.3: Steady-State Delay Model vs. Time Dependent Formulae

5.3.2 Unsignalized Intersections

Delay functions for unsignalized intersections, whether major/minor priority intersections or all-way-stop controlled intersections, are much simpler then functions for signalized intersections. For major/minor priority intersections, formulae developed by Kimber and Hollis (1979) are adopted by a related study of an asymmetric route choice model. Although the delay function developed by Kimber and Hollis is presently being tested in the ADVANCE Network using the asymmetric static route choice model, the BPR function is temporarily used for estimating delays at major/minor priority intersections with the following form (Bureau of Public Roads, 1964):

$$t = t_0 \left[1 + 0.15 \left(\frac{v}{c} \right)^4 \right] \tag{5.2}$$

where t is the link travel time, v is the link flow and c is the capacity of the link at a specified level of service. One main advantage of using the BPR function as a delay model for priority intersections is that it is defined for all flow-to-capacity ratios. This property is essential if the function is to be used in the context of a network equilibrium model. On the other hand, a significant drawback is that the BPR function is relatively flat at low v/c ratios, since it was originally conceived as a link performance function (Meneguzzer et al., 1990). Although using the BPR function as an intersection delay model is somewhat outside its designated scope, the lack of alternative functions with the desired analytical properties for use in a network equilibrium model mandates the use of the BPR function. Note that using the BPR function requires a non-zero free-flow travel time. Since intersection links in this research are coded as having a short length, it is necessary to attribute to them the physical characteristics (such as length and free-flow travel time) of their parent links, that is non-intersection links from where they originate.

As for the delay function for all-way-stop intersections, the following exponential delay model is used (Meneguzzer et al., 1990).

$$d = \exp \left[3.802(v/c) \right] \tag{5.3}$$

where d is the average approach delay (second/vehicle), v is the total approach flow and c is the approach capacity. Note the form of this exponential function, which is relatively flat at low flow-to-capacity ratio but becomes very steep as the degree of saturation increases, reflects the operational characteristics of all-way-stop intersections well. Kyte and Marek (1989) found that approach delay is approximately constant and in a range of five to ten seconds per vehicle for approach flows up to 300 to 400 vehicles/hour, but increases exponentially beyond this threshold. An increase in conflicting and opposing flows has the

effect of reducing this threshold. Equation (5.3) is suitable for use in a network equilibrium model, since it is defined for any flow-to-capacity ratio.

5.3.3 Freeway-Related Facilities

Several types of freeway-related facilities are found within the network including basic freeway segments, ramp junctions, weaving areas, toll plazas and ramps which are characterized as follows (Berka et al., 1994).

1. *Basic freeway segments* are those segments of the freeway not affected by merging or diverging movements at nearby ramps or by weaving segments.

2. *Ramp junctions* are the points at which on– and off–ramps join the freeway. The area around these points is in a state of turbulence due to the concentration of merging or diverging vehicles; for the purpose of this project, only delays due to on-ramps are considered.

3. *Weaving areas* are those segments of the freeway where two or more vehicle flows cross along a freeway segment. The freeway section between on– and off–ramps is considered a weaving area if an additional lane is provided and the distance between the ramp intersections is shorter than 3,200 feet.

4. *Toll plazas* are fully controlled access roadways with toll gates/booths for the purpose of collecting tolls from motorists.

5. *Ramps* are roadways designed to connect the arterial system with the freeway system; the on–ramps allow vehicles entering the freeway to merge smoothly with the through traffic on the freeway; the off–ramps allow the vehicles to exit the freeway and enter the arterial/collector road system

For the purpose of travel time calculations, four different freeway segments and one ramp segment are defined:

1. basic freeway segment;

2. basic freeway segment with an on– or off–ramp junction at downstream end of segment;

3. weaving area;

4. toll plaza;

5. ramp segment.

Table 5.3: Total Vehicle Flow per Hour

	Time Period	Total Flow
1	Night	19,439
2	AM Peak	184,185
3	Mid-day	170,573
4	PM Peak	203,278
5	Evening	146,092

In general, the travel time of the freeway segment is modeled as the travel time along the basic freeway section plus the delay at the bottleneck associated with the segment. The bottleneck is a location on the freeway where the vehicle flow is slowed or disturbed in some other way, causing deterioration of the travel condition upstream of that location. Ramp intersections and toll plazas are examples of bottlenecks.

5.4 Travel Demand and Time-Dependent Departure Rates

The ADVANCE Network is divided into 447 zones, originally specified by the Chicago Area Transportation Study (CATS), to assign time-dependent travel demand. Actually, these zones define a somewhat larger area called the extended test area. Daily trip tables based on CATS estimates for 1990 were factored to represent travel demand for five time-of-day periods that are listed below; see Zhang et al. (1994) for details.

Night	12 am to 6 am
Morning Peak	6 am to 9 am
Mid-day	9 am to 4 pm
Afternoon Peak	4 pm to 6 pm
Evening	6 pm to 12 am

The total vehicle flow per hour for the five time-of-day periods of the ADVANCE Network are shown in Table 5.3.

Each time-of-day period needs to be further divided into shorter time interval to solve the proposed dynamic route choice model. The 10–minute time interval is selected for applying in the solution algorithm. To load the travel demand represented by the O-D matrix onto the network time-dependently, 10–minute departure rates for each origin zone are derived from time-of-day half-hour departure rates obtained from CATS.

Steps used to derive the 10-minute flow departure rates are described as follows.

Step 1. Compute the normalized (sum to 1) half-hour flow departure rates of the analysis period (3-hour for the morning peak period and 2-hour for the afternoon peak period) based on the half-hour departure rates of 24-hour period obtained from CATS.

Step 2. Solve a nonlinear multiple regression model based on the half-hour flow departure rates of the analysis period obtained in Step 1. The selected nonlinear multiple regression model has the following form:

$$y = a_0 + a_1 x + a_2 x^2 + a_2 x^3 + \varepsilon \qquad (5.4)$$

where y is the flow departure rate (dependent variable), a_0 is the constant of this regression model, x represent the time interval (independent variable), a_1, a_2 and a_3 are the regression coefficients, and ε is the error variable.

Step 3. Substitute the values of the independent variables that represent the 10-minute time intervals into Equation (5.4) with estimates $\hat{a}_0$, $\hat{a}_1$, $\hat{a}_2$ and $\hat{a}_3$ obtained in Step 2. Then, estimates of the corresponding $\hat{y}$ are obtained.

Step 4. The 10-minute flow departure rates are computed by normalizing the estimates of the dependent variable ($\hat{y}$) obtained in Step 3.

Figures 5.4 and 5.5 show the time-dependent flow departure rates of the ADVANCE Network in the morning peak and afternoon peak, respectively; the flow departure rates are identical for all zones in the model.

5.5   Traffic Input Data

In order to calculate the delay (travel time) at diversified types of traffic facilities in the ADVANCE Network, besides the typical network supply data such as link capacity (or saturation flow) and free-flow travel time, additional traffic-related data are needed for this purpose including:

1. types of turning movements (e.g., left, through, right);

2. link facility types (e.g., centroid connector, freeway, tollway, arterial);

3. types of traffic control at intersections (e.g., signalized, priority, all-way-stop);

4. signal timing information (e.g., cycle length, green split).

Moreover, for the purpose of intersection delay calculations, collector and arterial intersections are further classified into 12 categories according to the type of *intersection control,*
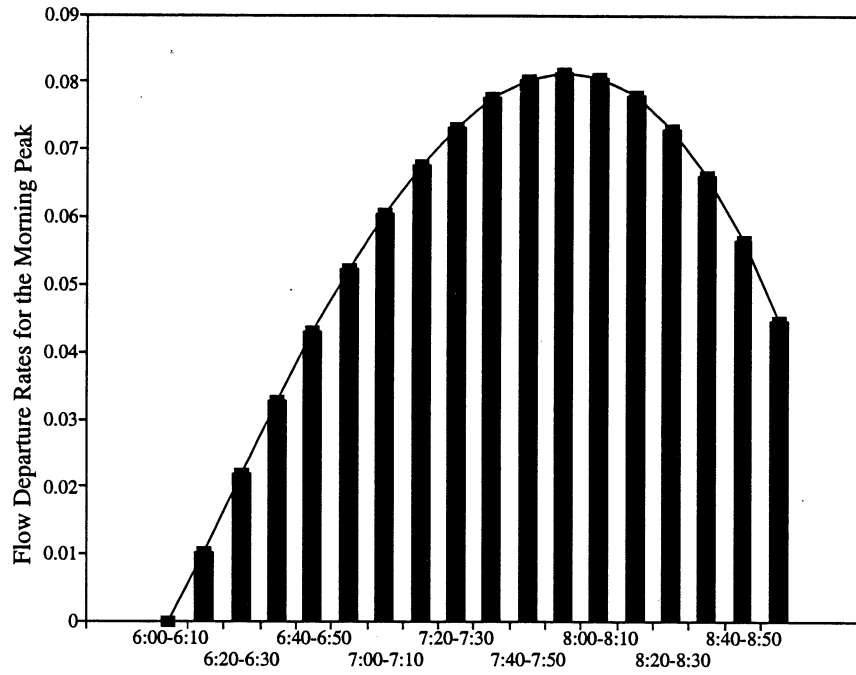
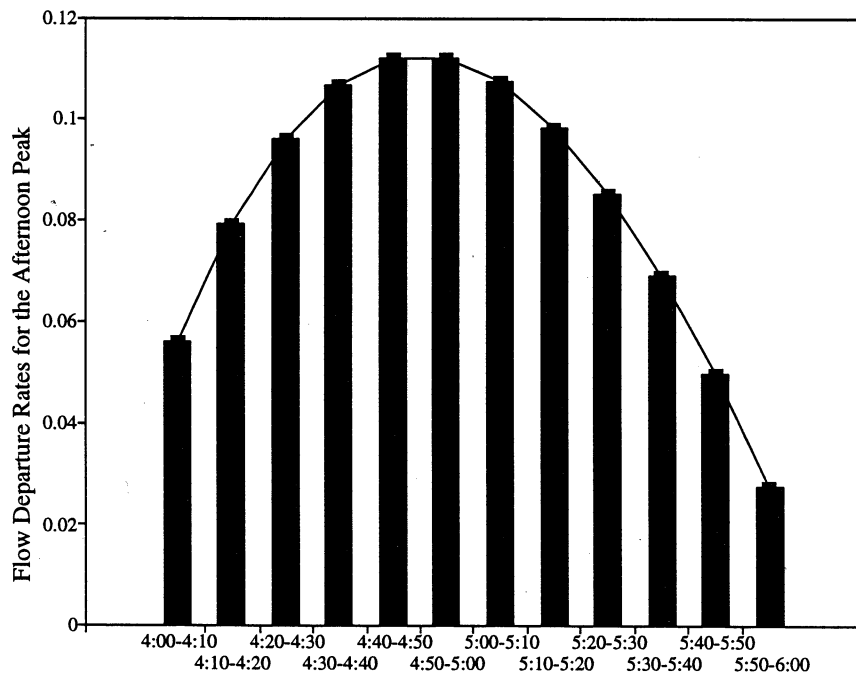Figure 5.4: Ten–minute Flow Departure Rates for the Morning Peak Period

Figure 5.5: Ten–minute Flow Departure Rates for the Afternoon Peak Period
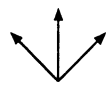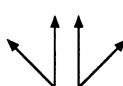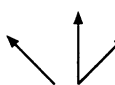
50
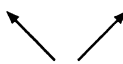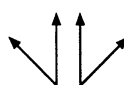
CATEGORY	CATEGORY NUMBER USED IN THE CODE	NUMBER OF LEGS	LANE DESIGNATION	CONTROL/PHASING
1	1	3 or 4		Signal, single phase
2	3	3 or 4		Signal, single phase
3	4	3 or 4		Signal, single phase
4	5	4		Signal, exclusive LT phase + all-movement phase
5	6	4		Signal, exclusive LT phase + all-movement phase
6	7	4		Signal, excl. LT and RT phase + all-movement phase
7	9	3		Signal, single phase
8	10	4	Unspecified	All-way-stop
9	11	3	Unspecified	All-way-stop
10	12	4	Unspecified	Major-minor priority
11	13	3	Unspecified	Major/minor priority
12	15	6 or 8		Signal, single phase

Figure 5.6: Classification of Street Approaches to an Intersection

51

Table 5.4: Intersection Approaches Classified by Category in the ADVANCE Network

Category	Number of Approaches
1	381
2	17
3	871
4	944
5	319
6	264
7	153
8	224
9	165
10	160
11	535
12	29
Total	4,062

the *intersection layout* and the *approach geometry*. Intersection categories are shown in Figure 5.6. Table 5.4 shows a breakdown of intersection approaches among these categories.

A few remarks about the classification of intersection categories are appropriate. Category 1 is used for any 1–lane approach. Category 7 is used only for 2–lane approaches when the opposing approach does not exist. The lane designation for unsignalized intersections (categories 8 through 11) is not specified because the delay functions adopted for these intersections do *not* depend on the geometry of the approach. The last category is used for all intersections with more than 4 legs, assuming only that each approach has an opposing approach. The above described classification procedure was originally developed by Meneguzzer et al. (1990) and revised by Berka et al. (1994) for the ADVANCE Network.

5.6 Alternative Route Choice Strategies

The problematic dynamic user-optimal assumption that complicates the modeling of many route choice options is that route choice decisions must be based on travel times which are temporally-consistent with future link flows at time of link use. The assumption is appropriate for recurrent trips and traffic conditions and is also acceptable for scheduled events (e.g., ballgames, concerts, detours, road constructions) and even for predicted weather conditions. However, this behavioral assumption is inconsistent with *unexpected* capacity reducing events (e.g., stalled vehicles, dropped objects, accidents) at future times because

drivers have very limited capability to be informed about the times and locations of such events before they encounter unusual queuing delays caused by those incidents. Enroute diversions are thus expected to occur when incidents are encountered by drivers.

5.6.1 Modeling Issues

The capacity reducing (lane blocking) events raise interesting questions with regard to how drivers make route choices. Two critical questions are (1) when do drivers learn of the blockage, and (2) how do drivers react to this knowledge. The three following route choice strategies may be very difficult (perhaps infeasible mathematically) to include within a mathematically valid and temporally-consistent dynamic route choice model:

1. time-of-departure route choice based only on current information;

2. enroute diversion when unusual traffic conditions are encountered;

3. enroute diversion guided by an in-vehicle information system.

Note all three alternative route choice strategies inevitably produce unequal-cost routes used by trips between any given O-D pair. Although this difficulty may seem minor (given that equilibrium solution is never reached exactly), a solution algorithm that is not designed to achieve the equal-cost O-D route objective can create erratic link flow fluctuations in successive time intervals (Janson and Robles, 1995).

The difficulty of modeling these three alternative route strategies is that equilibrium models use link travel time functions describing the route choice objective (e.g., *ideal* DUO) of all travelers on a given link in a given time interval. The adopted link travel time function is assumed to be known to each traveler at the time of route choice and is independent of trip origin and travel time to the link. Strategy (1) is clearly dependent on departure time and not the time interval of link is being used. Strategy (2) must assume anticipatory knowledge of traffic conditions beyond the bottleneck, or adopt the invalid assumption of strategy (1). That is, the remaining part of route is based on current knowledge only. Strategy (3) requires a very burdensome algorithm to update and revise routes in each time interval. All three strategies assume less than full anticipatory knowledge at the time of flow departure.

5.6.2 Modeling Approach

Considering the issues described in Section 5.6.1, there are two approaches that are mathematically valid and limit the possible impacts between the least and most severe cases.

1. The first approach is to assume zero incident diversion such that drivers always select expected routes based on anticipated traffic conditions without incidents. This

approach requires a set of alternative routes that is generated under usual traffic conditions and in a temporally-correct manner; then the routes will remain the same for flows even as routes are affected by incident-related queues and caused capacity reductions. This approach is similar to typical queuing analysis in which no route diversion is considered. This approach is anticipatory to usual traffic conditions but not anticipatory or reactive to unusual developing conditions. In general, this approach produces more severe queuing on incident routes than in reality because diversions are expected to occur under incident conditions.

2. The second approach is to assume that routes are selected according to full anticipatory knowledge. Consequently, this approach tends to underestimate the potential delays that are caused by incidents. However, the solution properties of this approach are quite clear and thus deficiencies of this solution are well understood. This approach can be extended as: one group of drivers choose routes based on time-of-departure traffic conditions, and another group of drivers diverts enroute because of incidents. In either case, full anticipatory knowledge remains the underlying assumption. This approach may produce less severe queuing on incident routes than in reality because excess diversions may occur because drivers are assumed to have too much information.

Despite violating the temporal consistency to some extent, the following method is employed to estimate the impacts of different percentages of drivers diverting enroute to alternative routes.

Step 1. First, assign the entire O-D matrix to the network assuming usual (incident-free) conditions. Save the aggregate link flows for each link in each time interval; a set of base-level link flows are thus generated. Denote this base-level array as *bvol*, where $bvol(a,t)$ defines the base-level flow on link a in time interval t.

Step 2. Next, select the percentage (say π) of all drivers that may change routes to avoid unusual congestions. Therefore, $(1-\pi)$ is the percent of all drivers that will *not* divert to alternative routes. That is, $(1-\pi)$ percent of drivers remain on the routes generated in Step 1 even when unusual congestion is encountered.

Step 3. Last, assign π percent of O-D matrix to the network and activate the links and time intervals that are affected by the capacity reduction events to estimate the flow (say $xvol(a,t)$) on each link in each time interval that may have rerouted. Consequently, the flow-to-capacity ratio must be computed on the basis of:

$$\text{flow-to-capacity ratio} = [xvol(a,t) + (1-\pi) \times bvol(a,t)]/capacity(a,t) \qquad (5.5)$$

Through this approach, drivers who choose the non-diverting strategy are actually based on fully anticipatory travel times when performing route choices. This approach might violate temporal continuity of flows in those time intervals where non-diverting base-level flows using the links are affected by enroute diversion flows. However, it becomes less severe as the non-diverting portion of drivers increases. Further, this approach also provides an estimate of related traffic characteristics when enroute diversions are guided by in-vehicle information.

Chapter 6

Computational Solution and Analysis of Results

In this chapter we present computational results for the link-time-based VI model of DUO route choice described in previous chapters. Five global network performance measures are defined to monitor the solution process of the model and assess the dynamic traffic condition over the ADVANCE Network. The selected global network performance measures are *average travel time, average travel distance, network space mean speed, average flow-to-capacity ratio* and *algorithm convergence index*. Next, we describe the platforms used for solving the model and their computing performance.

6.1 Computing Platforms and Performance

The algorithm was programmed in Fortran 77 and can be implemented on most available computing platforms including personal computers, workstations and supercomputers. For implementation on the ADVANCE Network, the model was executed on the CONVEX–C3880 at the National Center for Supercomputing Applications (NCSA), University of Illinois at Urbana-Champaign. The Convex–C3880 is a vector shared memory machine consisting of 8 processors (240 MFLOPS per processor peak), 4 Gbytes of memory and 60 Gbytes of disk space. For the ADVANCE Network with morning peak of 3 hours divided into 18 ten-minute time intervals, travel demand (552,597 total flow) and 512 Mbytes of memory, nearly 60 CPU hours are needed to reach a very fine convergence from a zero flow initial solution.

A modified version of the computer code can be executed with a much smaller memory requirement (55 Mbytes for the same problem size described above), and thus can be implemented on a workstation or even on a personal computer. However, it requires more CPU time and more disk space. The computational time of the modified version on a Sun Sparc 2 workstation (with 128 Mbytes of memory) is nearly 168 CPU hours.

6.2 Dynamic Network Performance and Convergence Measures

Five global network performance measures are defined to monitor the solution process of the model and assess the dynamic traffic condition over the ADVANCE Network. Definitions of these global network performance measures are listed below:

1. *Average travel time*

$$\bar{c} = \frac{1}{R} \sum_a \sum_t c_a[t] x_a[t]$$

where $\bar{c}$ =   average travel time (minutes)

R =   total flow during the analysis period (trips/period)

$c_a[t]$ =   travel time on link a at time interval t (minutes)

$x_a[t]$ =   flow on link a at time interval t

2. *Average travel distance*

$$\bar{\ell} = \frac{1}{R} \sum_a \sum_t \ell_a x_a[t]$$

where $\bar{\ell}$ =   average travel distance (miles)

ℓ_a =   length of the link a (miles)

3. *Network space mean speed*

$$\bar{S} = \bar{\ell}/(\bar{c}/60)$$

where $\bar{S}$ $\doteq$ network space men speed (mph)

4. *Average flow-to-capacity ratio*

$$\bar{x} = \frac{1}{K}\frac{1}{T} \sum_a \sum_t \frac{x_a[t]}{C_a[t]} x_a[t]$$

where $\bar{x}$ =   average flow-to-capacity ratio

K =   number of links in the network

T =   number of time intervals

$C_a[t]$ =   capacity of link a at time interval t

5. *Convergence Index*

The convergence index monitors the change in node time intervals between two consecutive *outer* iterations. The solution algorithm terminates if the change in node time

57

intervals (NDIFFS) is less than $nodes \times zones \times intervals \times x$. This index indicates that only x changes of the total node time intervals are allowed for the last flow that departed from each zone over the total analysis period. A sensitivity analysis of the x-value might be needed. With the perfect convergence, the changes of node time intervals equal to zero.

6.3 Analysis of Network Performance Measures

Selected traffic characteristics from the final *outer* iteration of the solution for the ADVANCE Network and separate road classes using the morning peak and afternoon peak travel demand are tabulated in Tables 6.1 and 6.2. Figures 6.1, 6.3, 6.5 and 6.7 show the variations in the global network performance measures for morning peak solutions of the *outer* iterations of the algorithm. The results for the afternoon peak are shown in Figures 6.2, 6.4, 6.6 and 6.8.

For analytical-based route choice models, average travel times tend to decrease with the iterations of the algorithm, indicating flows are assigned to better routes leading to shorter travel times. Both Figures 6.1 and 6.2 show an overall pattern of decreasing average travel times. Based on the results shown in Tables 6.1 and 6.2, the afternoon peak is slightly more congested than the morning peak for the ADVANCE Network. In addition, arterials have higher travel times than collectors and freeways (see Figures 6.1 and 6.2), showing that about 70% of the average flow occurs on arterial links of the ADVANCE Network, as measured by travel time, and about 60% as measured by distance. Note in Figure 6.1 that the average travel times of collectors increase sharply in the fifth *outer* iteration; a slight increase in average travel times of arterials is observed in the same iteration. According to the computational experience of this model, the average travel times decrease in the first few iterations, and then oscillate before the defined fine convergence is reached. However, further investigation of the increase in average travel times may be revealing.

Figures 6.3 and 6.4 indicate that the average travel distance does not vary much among the solution iterations, implying the used routes are found quickly, but the algorithm still needs to adjust the *node time intervals* to reach the convergence. These results show that the afternoon peak period has a slightly shorter average travel distance than the morning peak in the ADVANCE Network. Considering the layout of freeways in the ADVANCE Network, the usage of freeways seems low based on the results shown in Table 6.1 and 6.2. In the ADVANCE Network, arterials are used extensively in the morning and afternoon peak.

A general observation is that the network space mean speeds tend to be increasing with the *outer* iterations, which is reasonable since the average travel times are decreasing. Intuitively, freeways have the highest space mean speed, then arterials, and finally collectors.

Table 6.1: Solution Characteristics for Morning Peak Period by Road Class

Link Class	Travel Distance (miles)	Travel Time (minutes)	Mean Speed (mph)
Collector	1.18 (1.34)	4.20 (5.97)	16.86 (13.47)
Arterial	6.03 (6.83)	17.42 (21.65)	20.77 (18.92)
Freeway	2.82 (3.01)	3.62 (3.95)	46.74 (45.73)
All Classes	10.02 (11.18)	25.24 (31.61)	23.81 (21.22)

(·) results from Berka et al. (1994)

Table 6.2: Solution Characteristics for Afternoon Peak Period by Road Class

Link Class	Travel Distance (miles)	Travel Time (minutes)	Mean Speed (mph)
Collector	1.16 (1.36)	4.04 (5.69)	17.23 (14.37)
Arterial	5.85 (6.60)	18.66 (22.78)	18.81 (17.37)
Freeway	2.79 (3.11)	3.59 (4.11)	46.63 (45.37)
All Classes	9.80 (11.07)	26.29 (32.63)	22.37 (20.35)

(·) results from Berka et al. (1994)

Space mean speeds on arterials and freeways of the afternoon peak are lower than for the morning peak.

Figures 6.7 and 6.8 show changes in the flow-to-capacity ratios of the morning and afternoon peak periods, respectively. Since the DUO flow pattern is obtained after several *outer* iterations, it is reasonable to see the decreasing flow-to-capacity ratios. These results imply that the successive adjustments of the node time intervals lead to a DUO state. In the last *outer* iteration, the flow-to-capacity ratio is 0.68 for the morning peak versus 0.69 for the afternoon peak.
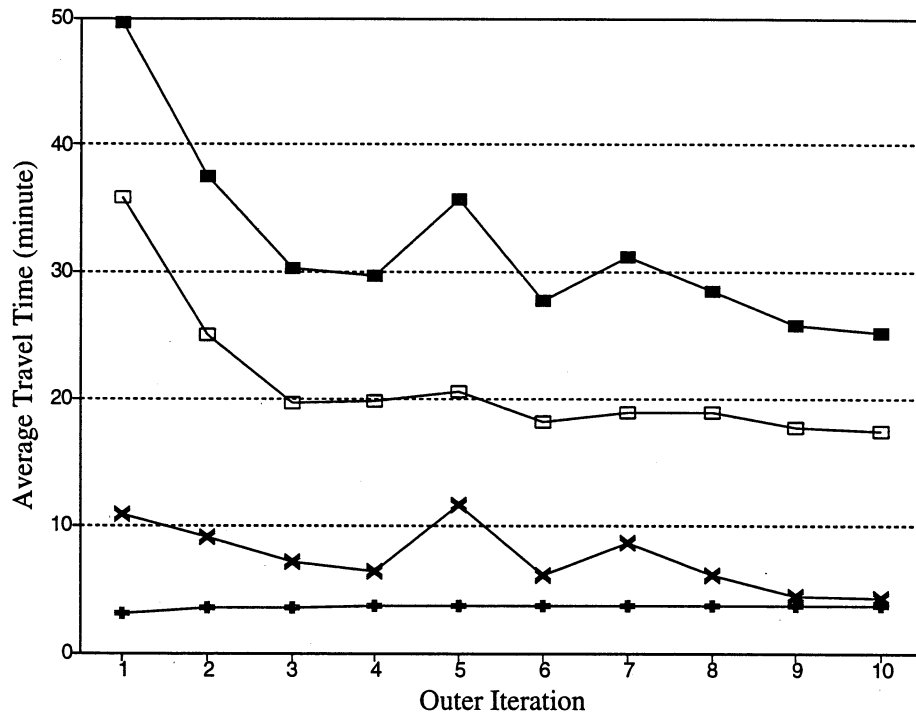
Figure 6.1: Average Travel Time of the Morning Peak Period

Compared with the results from the asymmetric SUO route choice model (Berka et al., 1994), the dynamic model of the ADVANCE Network exhibits less congestion based on the selected network performance measures described above. Besides generating these global network performance measures, the model is capable of providing time-dependent traffic-related information such as link travel time, travel speed, capacity and queue spillbacks.

Unfortunately, empirical link flow and link travel time data are not available for the ADVANCE Network, either in general, or more specifically for the O-D matrix used in this solution. These data, as well as route flow data, are urgently needed to advance the state of the art of network modeling for ITS. Although not fully validated, the results from this model are internally consistent.
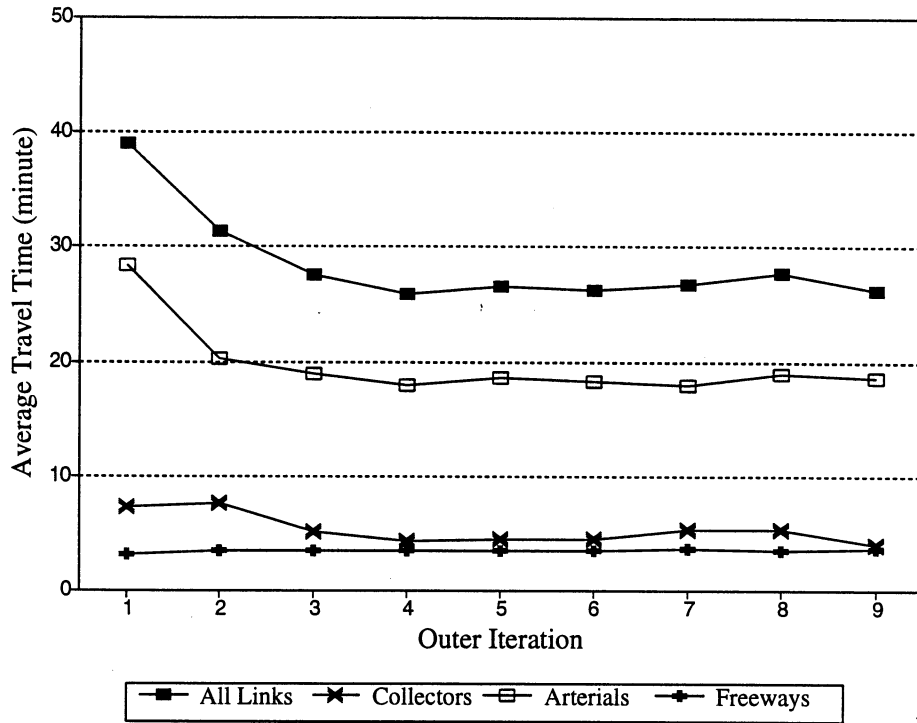
Figure 6.2: Average Travel Time of the Afternoon Peak Period
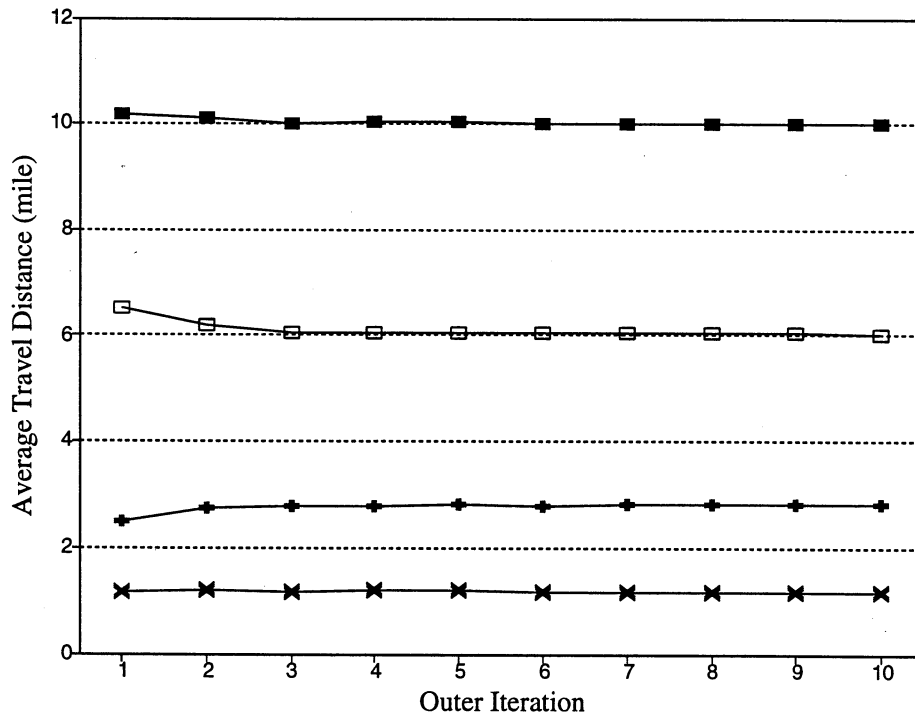
Figure 6.3: Average Travel Distance of the Morning Peak Period

61
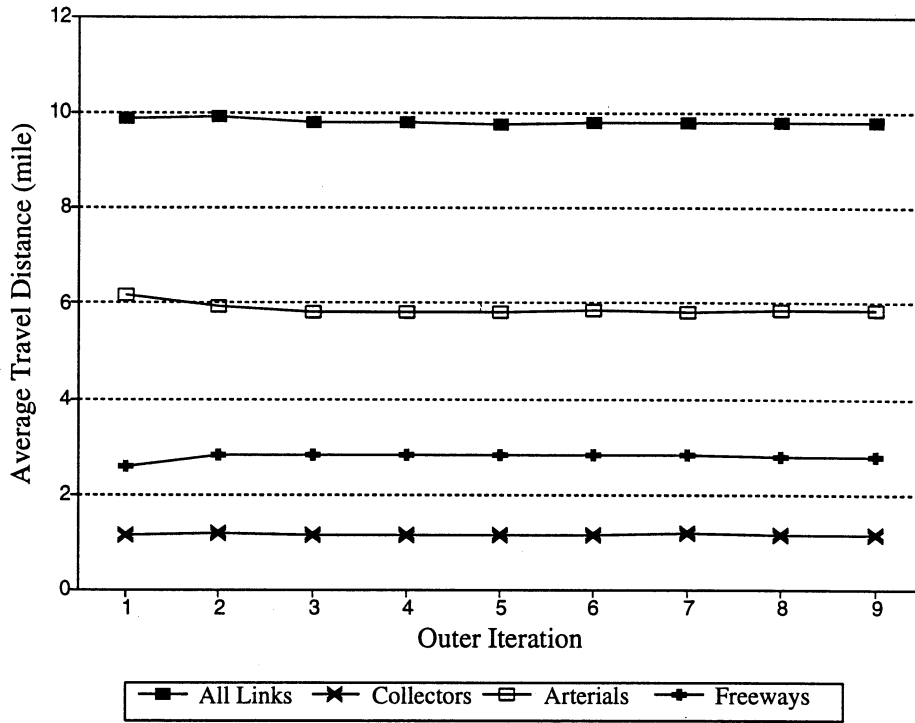
Figure 6.4: Average Travel Distance of the Afternoon Peak Period
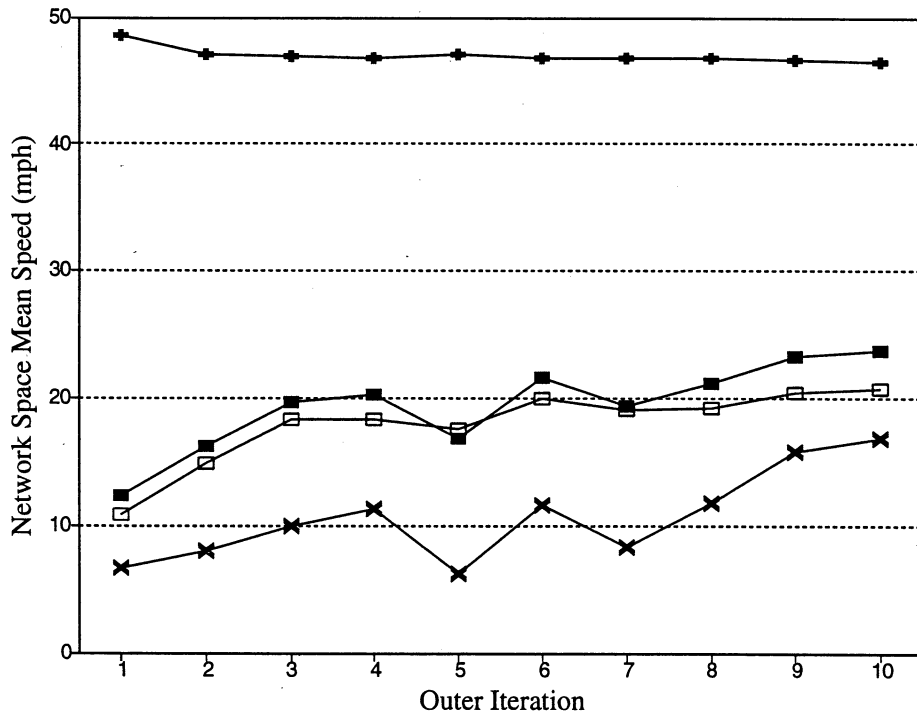
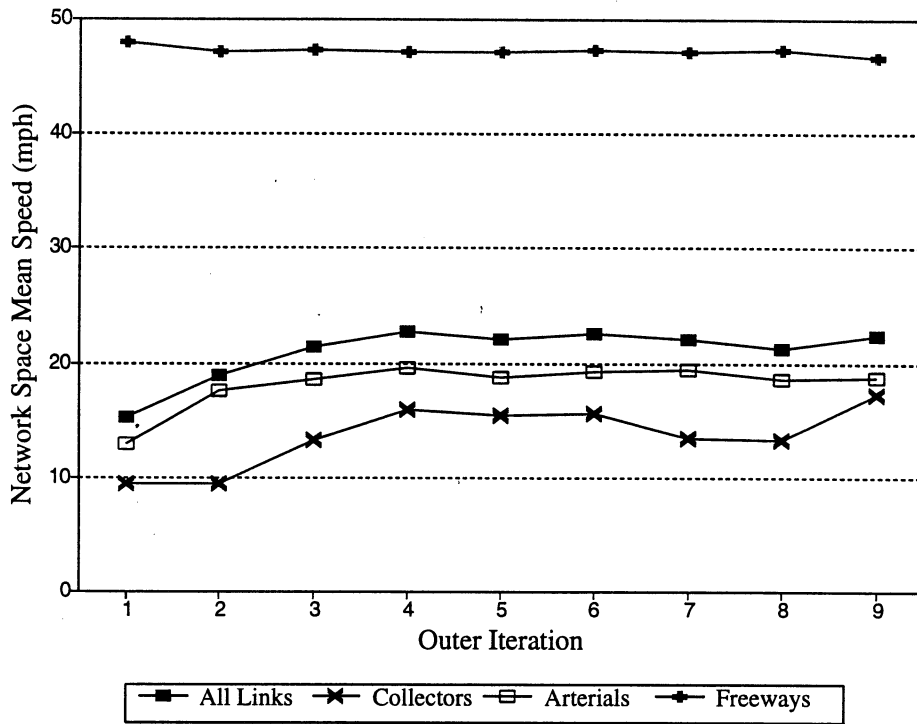Figure 6.5: Network Space Mean Speed of the Morning Peak Period

62

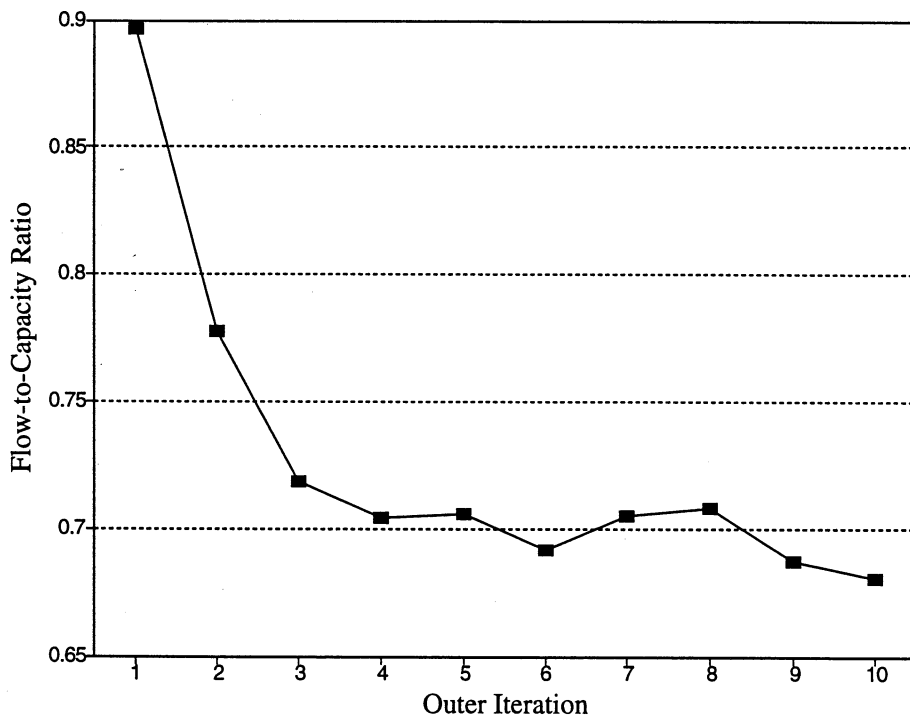Figure 6.6: Network Space Mean Speed of the Afternoon Peak Period

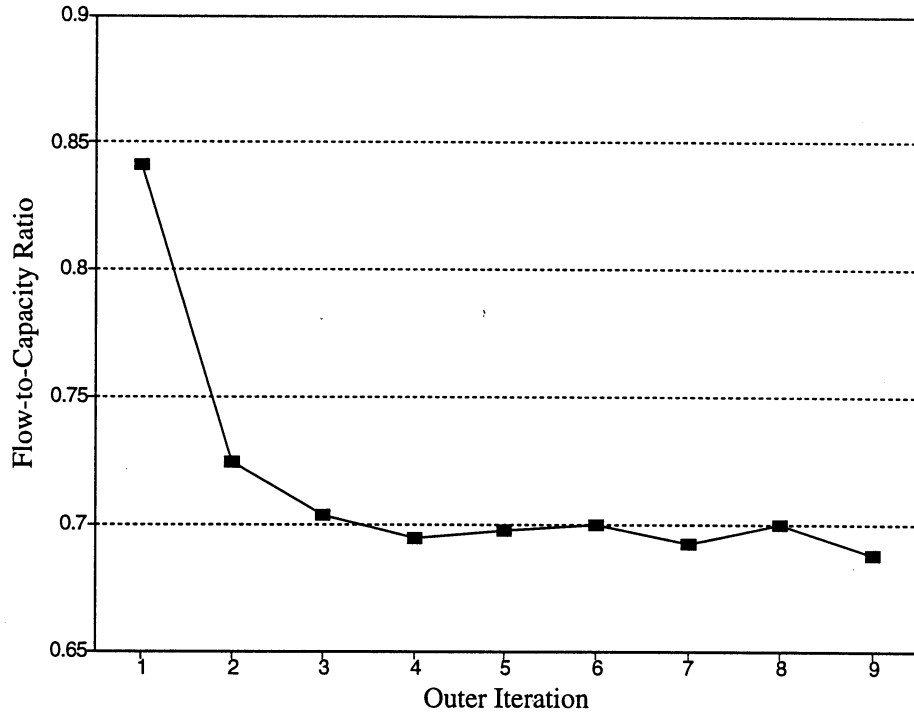Figure 6.7: Flow-to-Capacity Ratio of the Morning Peak Period

63

Figure 6.8: Flow-to-Capacity Ratio of the Afternoon Peak Period

For the convergence measures, *nodes* = 9,700, *zones* = 447, *intervals* = 18 for the morning period and 12 for the afternoon period; $x = 0.001$ is used in the solution algorithm. A rather small value is chosen for x, indicating a fine convergence of the algorithm is desired.

After 10 *outer* iterations, NDIFFS equals to 20,537 showing the algorithm has converged at the level of $x \approx 0.00026$ for the morning peak period. For the afternoon peak period, the algorithm converges after 9 *outer* iterations with NDIFFS equal to 25,658, indicating the algorithm indeed converges at the level of $x \approx 0.00049$.

As shown in Figures 6.9 and 6.10, the rate of change of the *node time intervals* is calculated as the change in the *node time intervals* divided by the total possible change of the *node time intervals* (i.e., *nodes* × *zones* × *intervals*; 78,046,200 for the morning peak period and 52,030,800 for the afternoon peak period) between consecutive *outer* iterations. Figures 6.9 and 6.10 display the rates of change of the *node time intervals* which indicate that this model was solved quite smoothly both for the morning and afternoon travel demand. From Figures 6.9 and 6.10, we can find the rate of change of the *node time intervals* moves faster in the first four iterations than in the later iterations.

If a higher value of x is chosen for calculating NDIFFS, the algorithm converges in only four *outer* iterations, yielding a saving of nearly 60% of CPU time. Based on the computational experience, $x = 0.006$ is an acceptable value for applications. Note that we
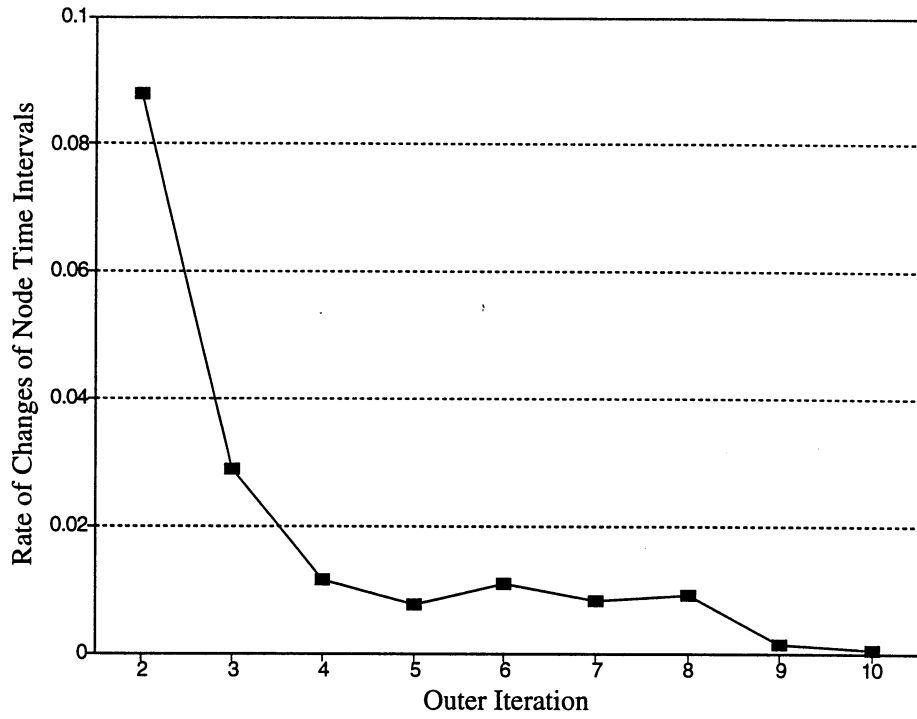
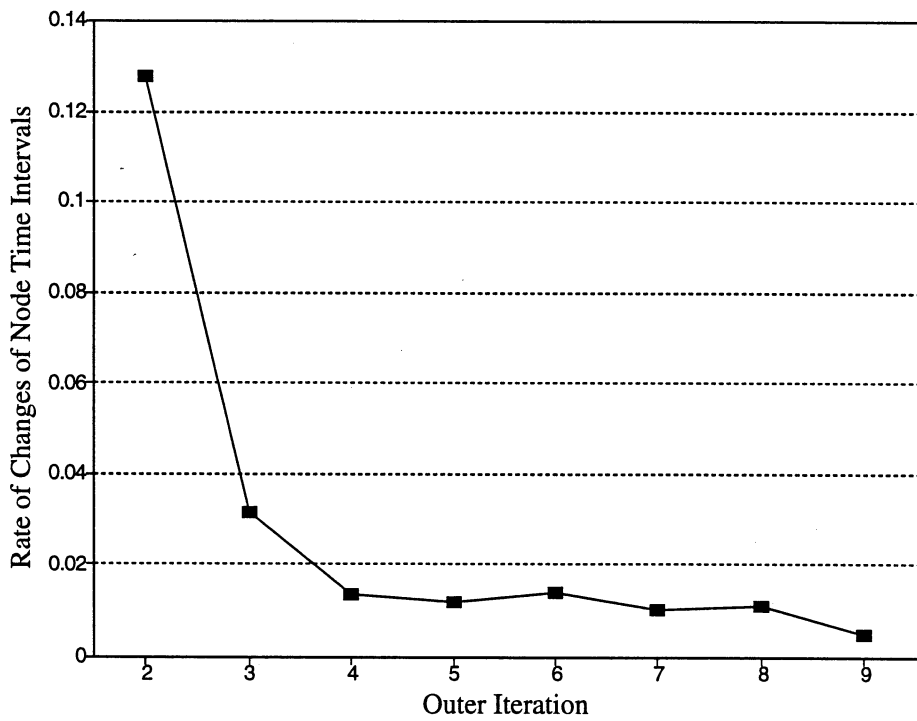Figure 6.9: Rate of Change of Node Time Intervals of the Morning Peak Period

Figure 6.10: Rate of Change of Node Time Intervals of the Afternoon Peak Period

omit the value of the first *outer* iteration in Figures 6.9 and 6.10 to provide a better display of the variations in the rest of iterations.

6.4 Enroute Diversions Resulting from Incidents

In this section, we present the results of an analysis of incidents. First, we place an incident on a through movement of an intersection on a southbound major arterial and assume that 50% of drivers choose a diversion strategy to avoid the incident. Next, we place another incident on a major eastbound arterial which is a primary traffic corridor in the southeast corner of the test area. We assume no driver chooses a diversion strategy since this vicinity offers no alternative routes in the incident direction (eastbound).

6.4.1 Case 1
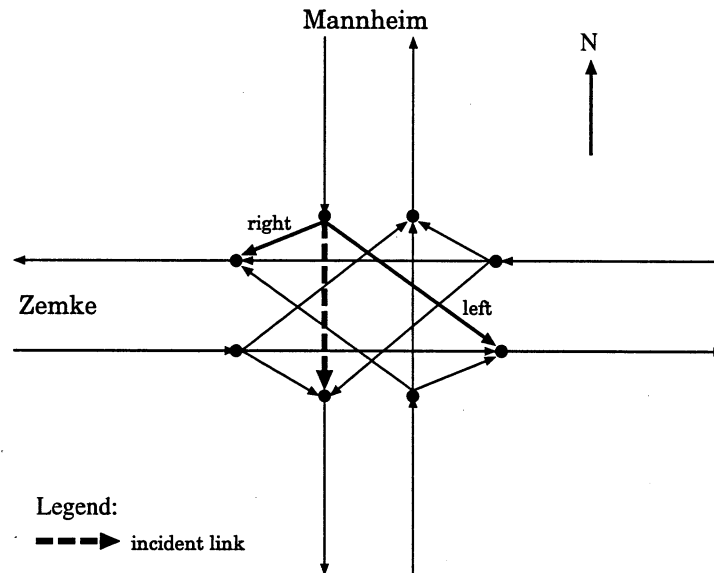
Figure 6.11: Layout of the Incident Analysis Area–Case 1

To demonstrate the effects of enroute diversions that result from incidents, an incident is placed on a southbound through-movement link of an intersection which causes a 50% capacity reduction from time interval 7 to 12 (7 AM to 8 AM), one-half of the drivers were assumed to choose a diversion strategy. The intersection is a four-leg signalized intersection on a major arterial with no turning restrictions (U-turn excluded). Note that the scale of capacity reduction can also be specified over intervals to account for the gradual loss of capacity caused by the incident. Figure 6.11 shows the layout of the incident intersection.

Figure 6.12 shows the flow profile of the incident link by time interval. Obviously, the flows on the incident link are lower than the non-incident flows during the incident period.

66

Following the removal of the incident, the incident flows become higher than the non-incident flows. Although the flows of the incident condition are lower than the non-incident condition, the link travel times of the incident condition are still higher than the non-incident condition because of the loss of capacity (see Figure 6.13). The upstream link has similar results as the incident link as shown in Figures 6.14 and 6.15. Since intersection links are being discussed, link travel speeds are not interesting to show in this case.
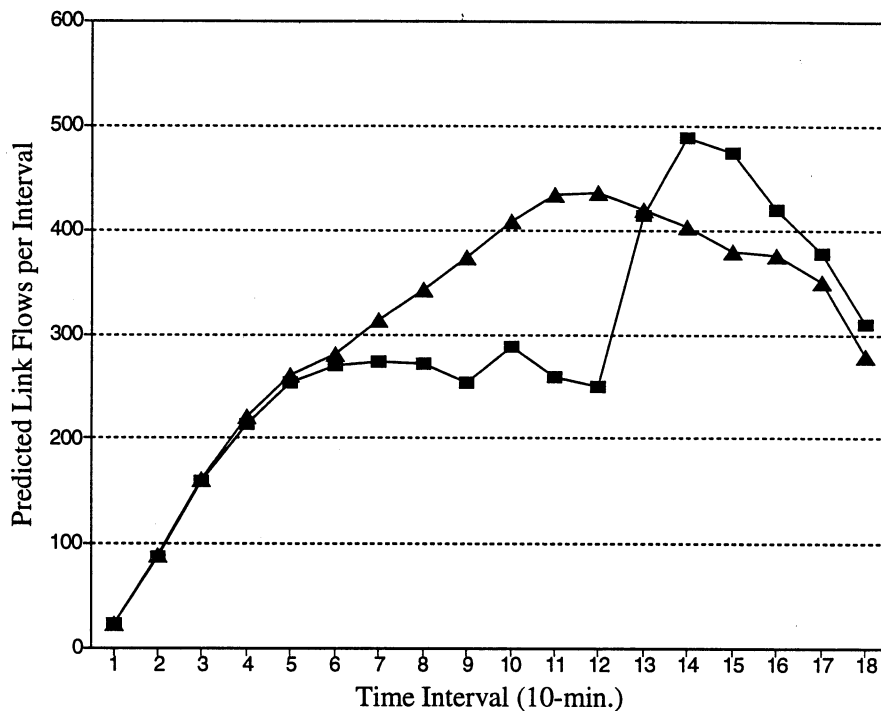
Figure 6.12: Predicted Flows of the Incident Link–Case 1

Significant flow diversions appear in the right-turn and left-turn links during the incident periods (see Figures 6.16 and 6.17). Figures 6.18 and 6.19 show the variations of travel times. The diversions on the right-turn link are quite obvious; the incident flows are higher than the non-incident flows during the incident period and a few subsequent time intervals.

The flow profile of the left-turn link is complex, but is still reasonable. The incident flows begin to exceed the non-incident flows at the second incident interval because of the normal left-turn delay. The incident flows drop soon and return to the usual used links following the clearance of the incident because of the capacity restoration on the incident link. Although diversion could occur in upstream intersections, results shown in Figures 6.16 and 6.17 imply flows are choosing alternative routes. Based on this modeling capability, this model can be utilized as a tool to assess an area-wide incident management strategy and the resulting flow pattern.
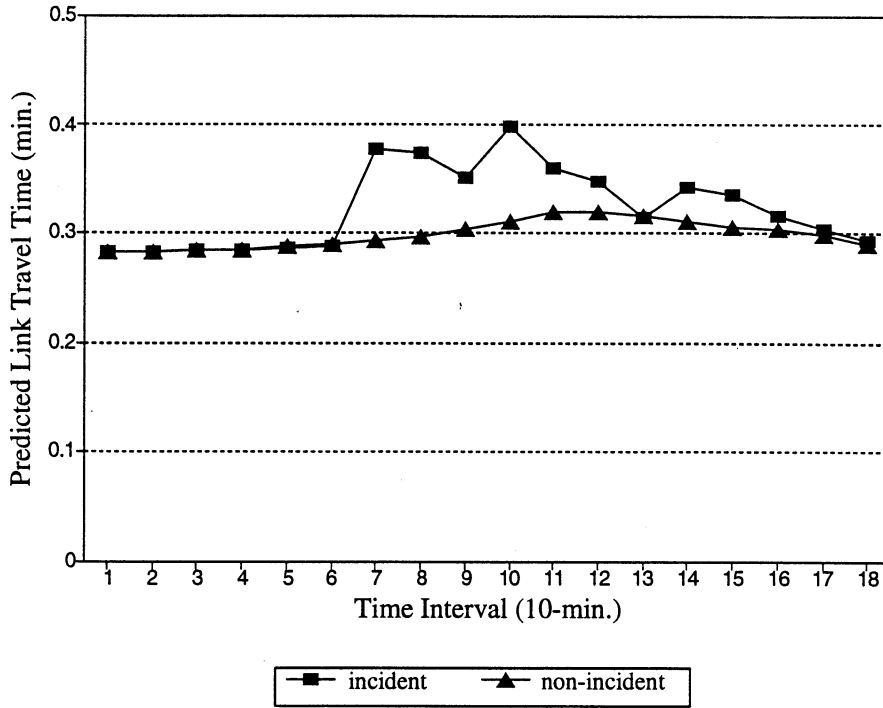
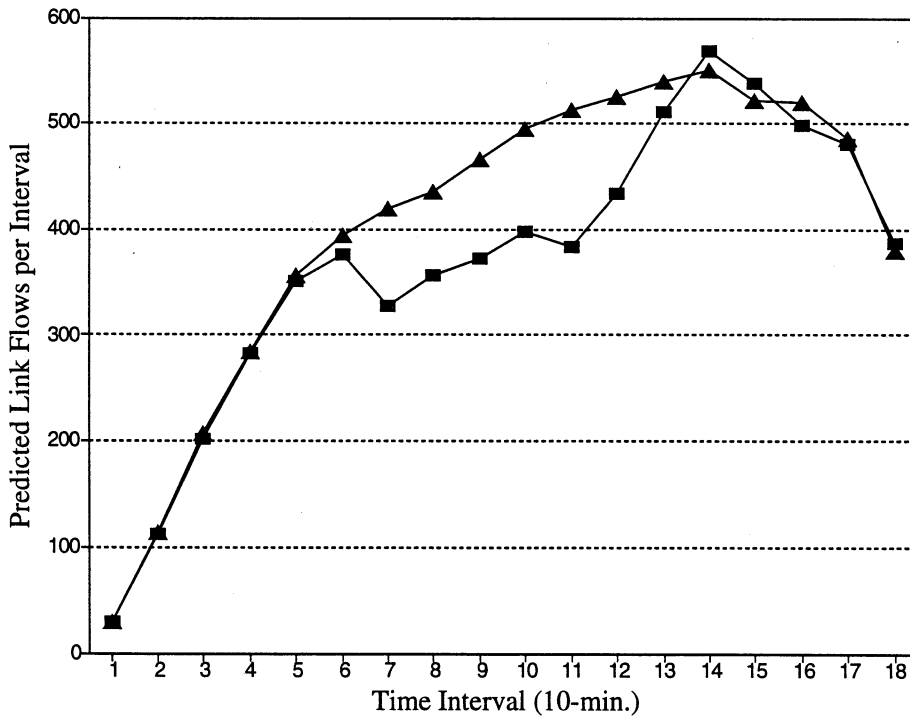Figure 6.13: Link Travel Times of the Incident Link–Case 1

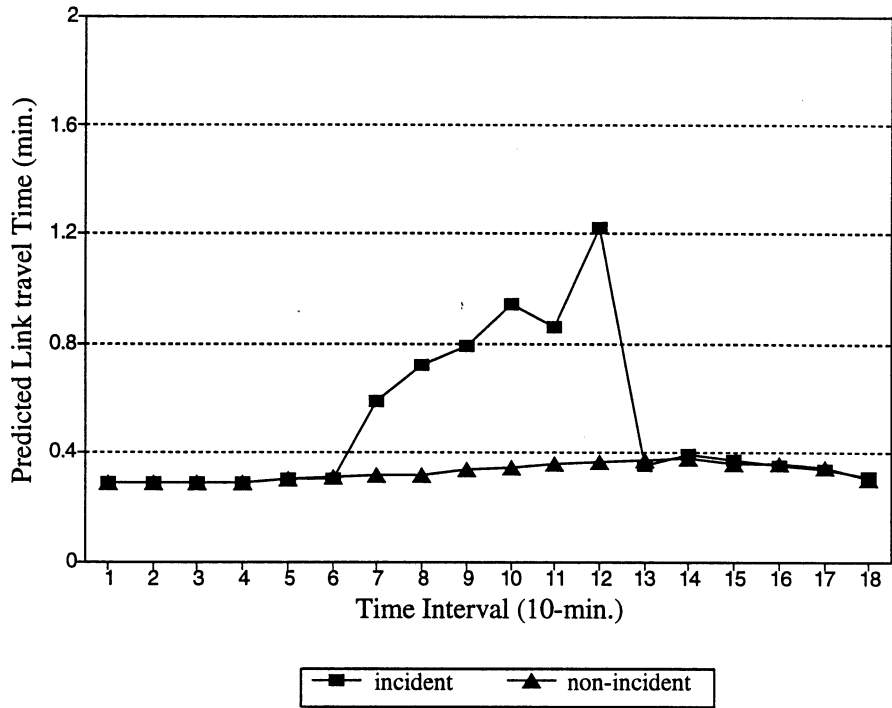Figure 6.14: Predicted Flows of the Upstream Link–Case 1

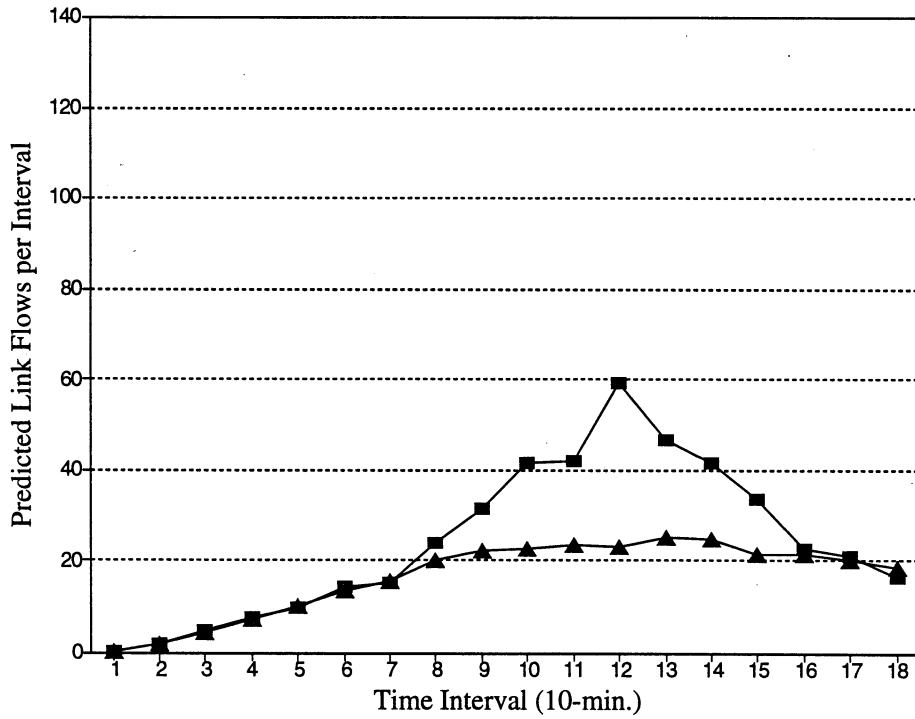Figure 6.15: Link Travel Times of the Upstream Link–Case 1

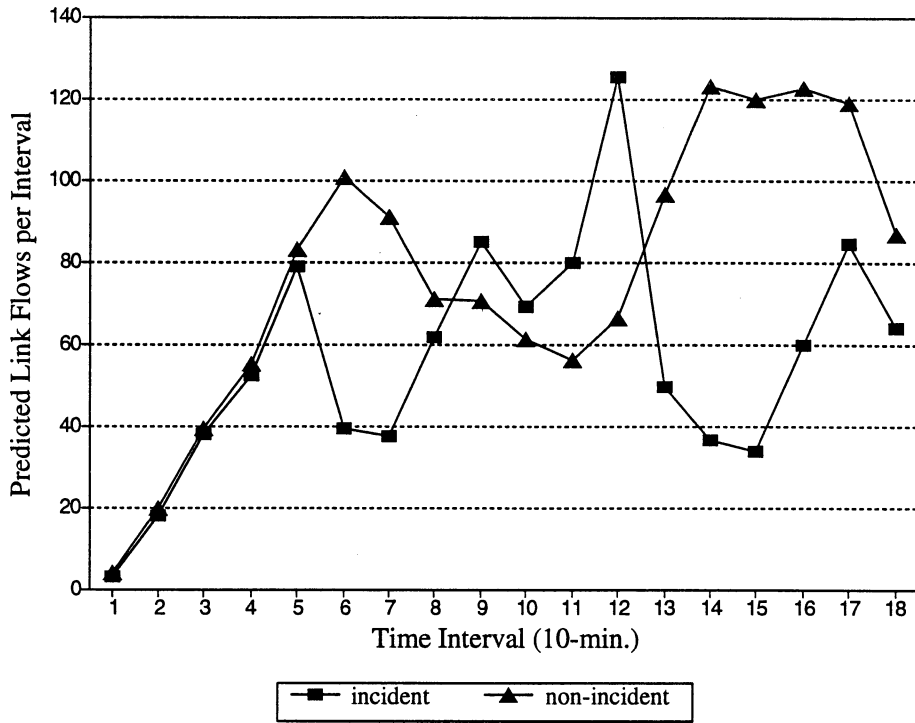Figure 6.16: Predicted Flows of the Right-Turn Movement–Case 1

69

Figure 6.17: Predicted Flows of the Left-Turn Movement–Case 1
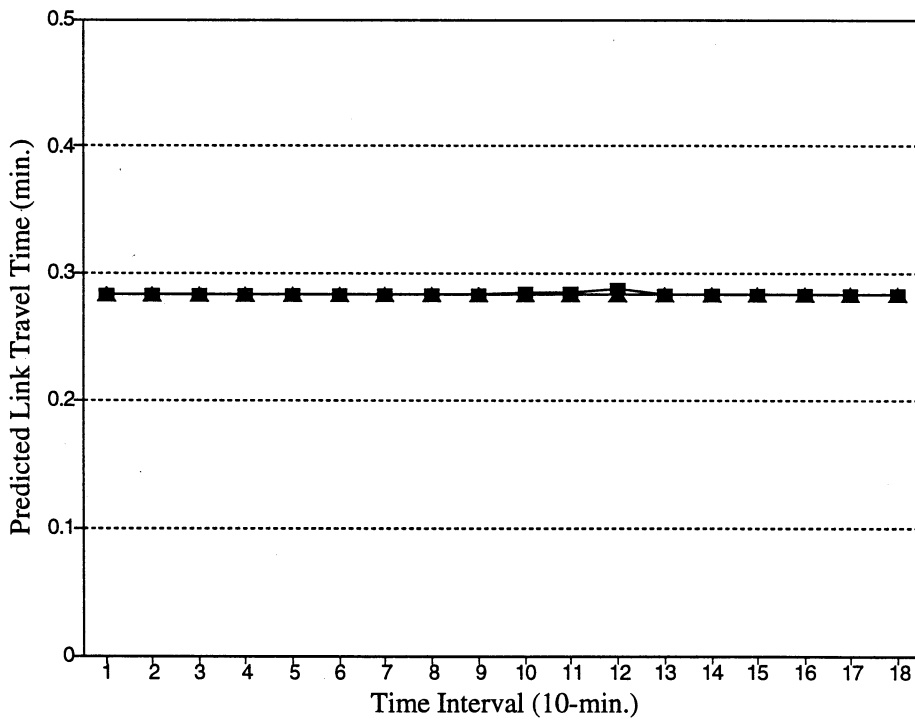

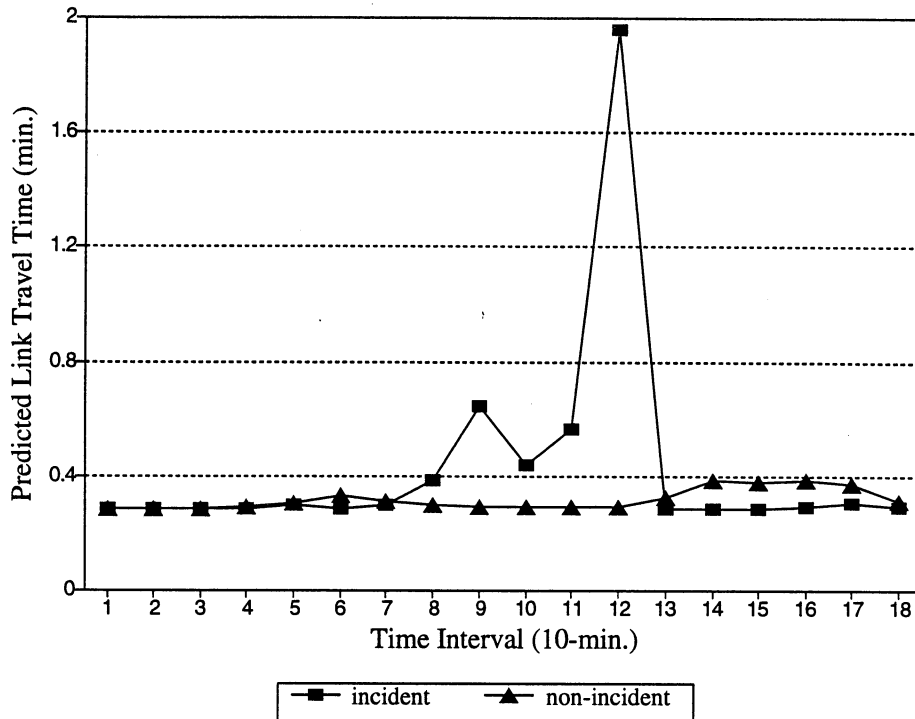Figure 6.18: Link Travel Times of the Right-Turn Movement–Case 1

Figure 6.19: Link Travel Times of the Left-Turn Movement–Case 1

6.4.2 Case 2

Now, a two car collision is placed on the inside lane of a two-lane eastbound arterial which is a major corridor in the vicinity. Figure 6.20 shows the layout of the analysis area. As mentioned previously, the scale of capacity reduction can be specified over intervals to account the gradual capacity loss caused by the incident. The following capacity reductions are applied:

Interval	Action	Capacity Loss
4:40–4:50 PM	accident occurs	65%
4:50–5:00 PM	police arrive	80%
5:00–5:10 PM	tow car arrives	85%
5:10–5:20 PM	accident is removed	65%

We assume a traffic accident occurs at 4:40 PM and persists for 40 minutes (i.e., from time interval 5 to 8). In the incident interval, the traffic accident occurs and blocks 65% of the link's capacity. The police come to the scene and block one lane to process the accident, which causes a loss of 80% of the link's capacity during the second incident interval. In the third incident interval, the involved cars are moved to the shoulder and tow cars arrive,

71

causing an 85% capacity reduction. In the fourth incident interval, the involved cars are towed away from the scene, and traffic returns to normal; 65% loss of capacity is assumed.
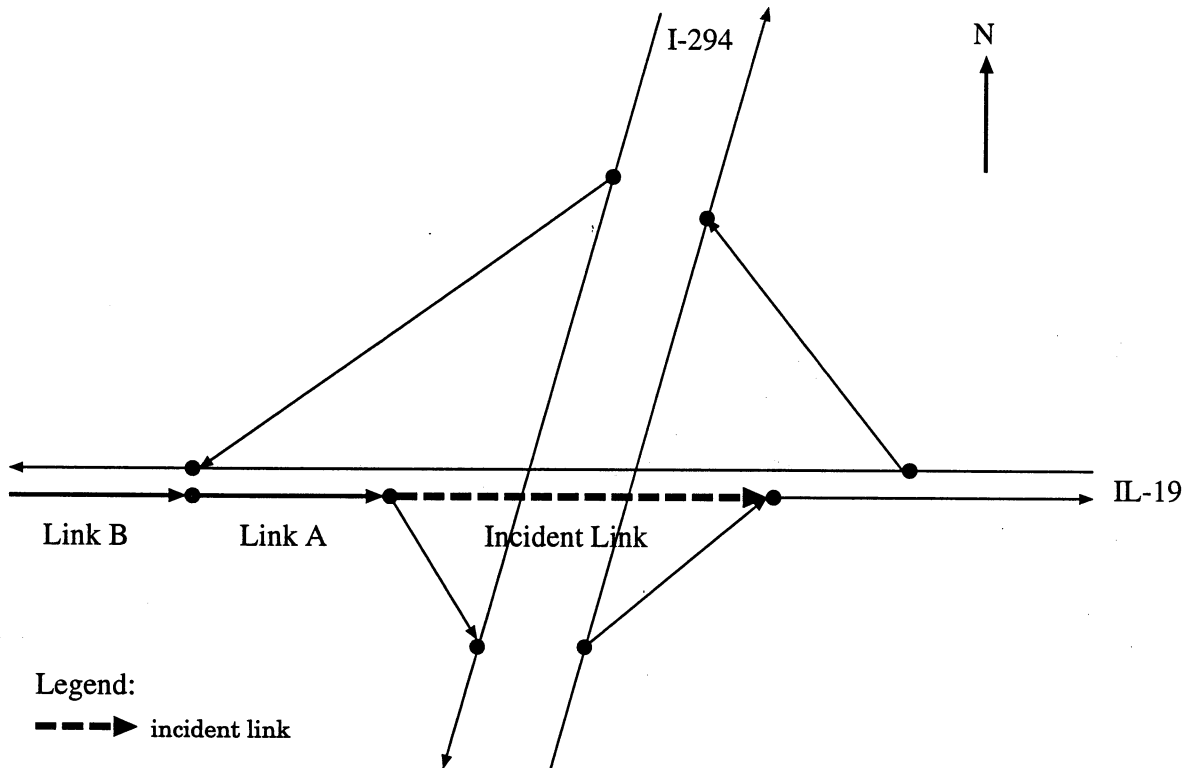

Figure 6.20: Layout of the Incident Analysis Area–Case 2

As shown in Figure 6.21, the link travel times for the incident condition during the incident period are significantly higher than for the non-incident condition. Because of the zero diversion assumption, the link flows remain the same both for the incident and non-incident conditions (Figure 6.22); therefore, flows are not delayed in this incident scenario. Consequently, the link travel time is about 13 minutes in the most congested time interval (7). This treatment will generate vertical queues on affected links so that queues will not propagate properly under this condition. Although this may be an unrealistic treatment and can be viewed as a shortcoming of this modeling approach, it provides a severe impact estimate of the incident.

The highest flow of the incident link appears in time interval 6 (see Figure 6.22), but the highest travel time occurs in time interval 7 due to the most severe capacity loss and queuing effect, captured by the link capacity adjustments of the solution algorithm. Although the incident period ends in time interval 8 (5:20 PM), the queuing effect results in the travel time returning to the normal condition only in time interval 10. Similar observations apply to the link travel speed (see Figure 6.23).
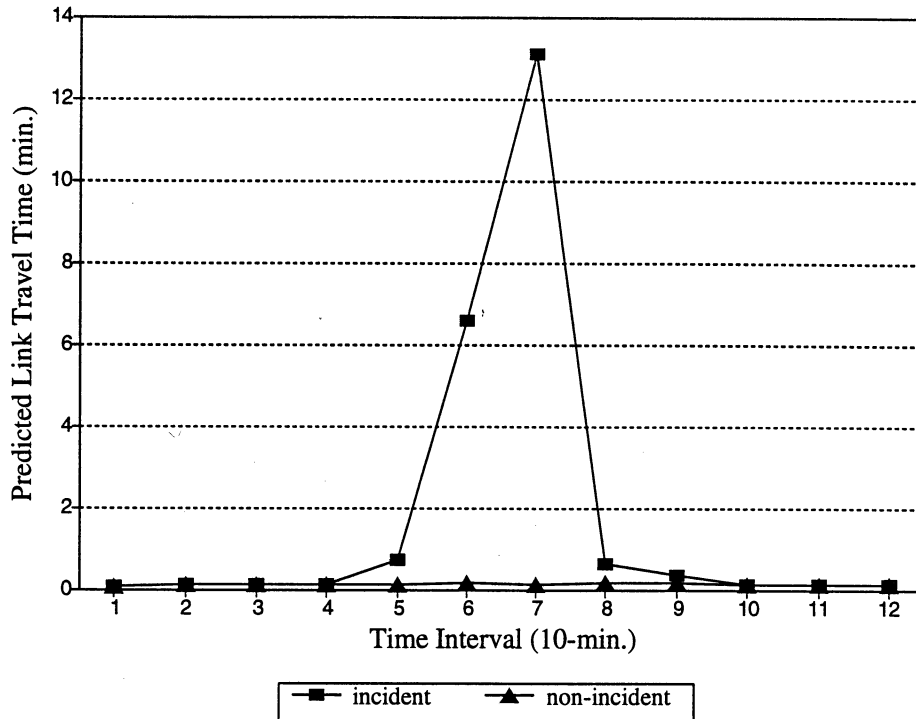
Figure 6.21: Link Travel Times for Incident and Non-Incident Conditions–Case 2

Since the vicinity offers no alternative routes in the incident direction, the solution algorithm detects that the incident causes queue spillbacks to the upstream links. As shown in Figure 6.20, links A and B are affected by the spillback queue originating from the incident link.
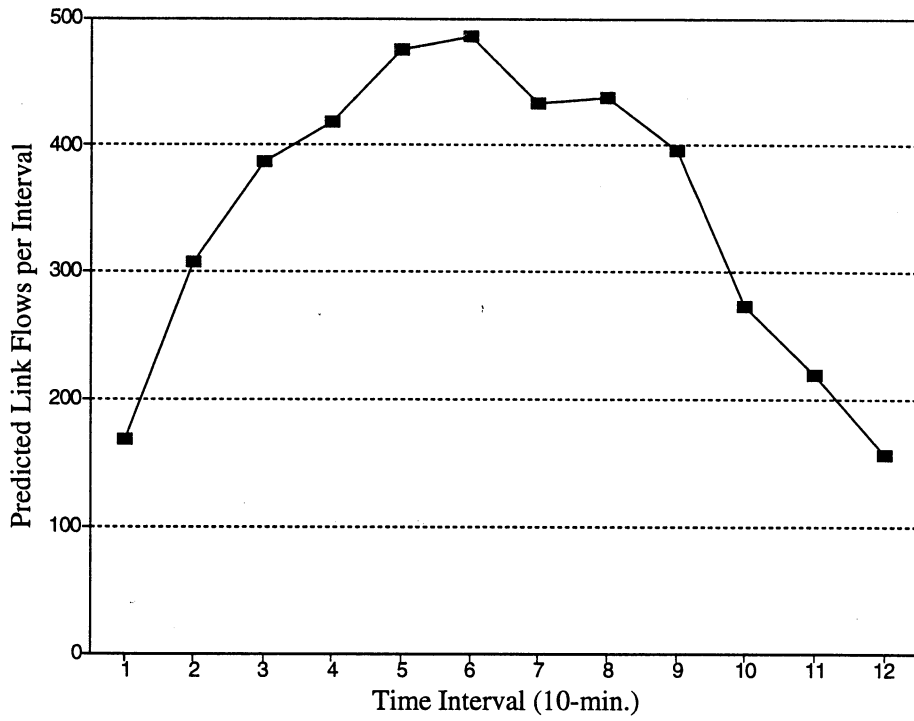
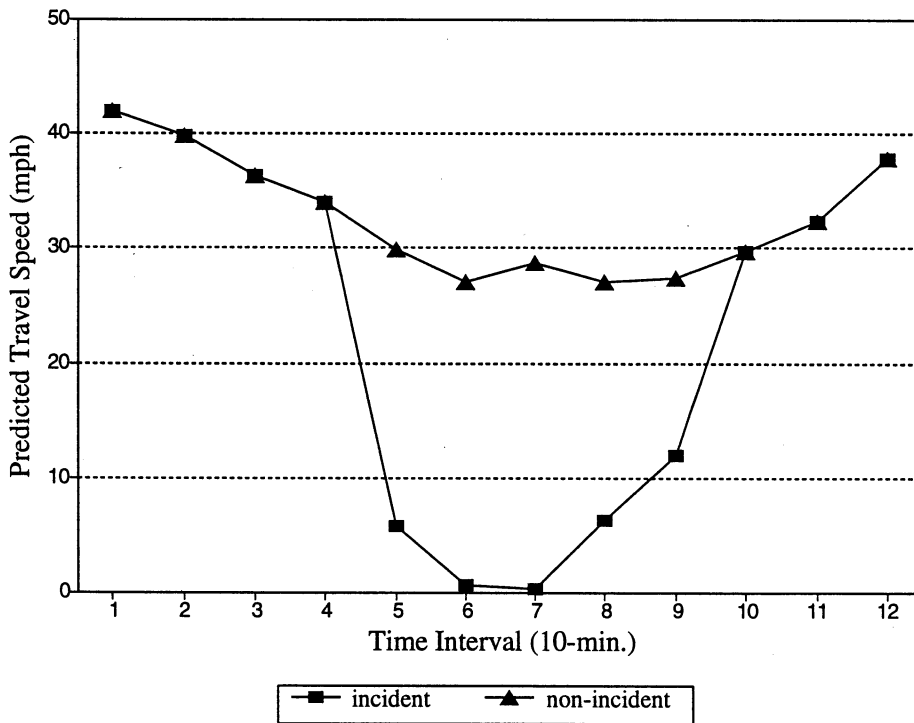Figure 6.22: Link Flows for Incident and Non-Incident Conditions–Case 2

Figure 6.23: Link Travel Speed for Incident and Non-Incident Conditions–Case 2

Chapter 7

Conclusions and Future Research

In this concluding chapter, we offer an account of the contributions and the notable findings of this research in Section 7.1. Recommendations for future research are presented in Section 7.2.

7.1 Conclusions

This research presented a quasi-continuous time formulation of the link-time-based variational inequality model of dynamic user-optimal route choice for evaluating time-dependent traffic characteristics for Advanced Traffic Management Systems (ATMS) and Advanced Traveler Information Systems (ATIS) such as the recently concluded ADVANCE Project.

The proposed DUO route choice model can be efficiently solved to convergence by using the diagonalization algorithm described. For most links a realistic traffic engineering-based link travel time function, the Akcelik function, is adopted in this research, in place of the simplistic but widely used BPR (Bureau of Public Roads) function, to estimate delays and travel times for various types of links and intersections. An expanded intersection representation mechanism is employed so that each turning movement is treated as an individual link. To this end, nearly 23,000 links and 10,000 nodes are modeled in the solution process. The time-varying total flows of the morning peak (6 AM to 9 AM; 552,597 trips per 3-hour period) and the afternoon peak (4 PM to 6 PM; 406,556 trips per 2-hour period) in the ADVANCE Network are solved by the algorithm.

Various time-dependent traffic characteristics such as link flow, travel time, travel speed and flow-to-capacity ratio are available both at the link and network-wide levels. Furthermore, queue spillback information is also generated from the model to detect unusual queues and congested bottlenecks in the network. The converged solution is obtained and analyzed. The problem scale solved in this research is believed as the first and the largest solution of an analytical-based dynamic route choice model obtained to date. In particular, this model

is solved for a real-world traffic network rather than a hypothetical network.

Unexpected lane-blocking events that cause nonrecurrent traffic congestion are capable of being analyzed with this model. Although full anticipatory knowledge of traffic conditions remains the underlying assumption, the approach described in Section 5.6.2 provides an estimate of the impact of different percentages of drivers diverting enroute to alternative routes. This approach provides an estimate of enroute diversions, if guided by an in-vehicle information system.

Although not yet fully validated, this model is able to predict time-dependent traffic characteristics for a large-scale traffic network which are reasonable and internally consistent. Eventually, dynamic route choice models should be integrated into a traffic management center to support the decisions on the adjustments of arterial signal timing, ramp metering, incident management and future route guidance strategies, etc.

Significant contributions of this research are identified as follows.

1. This research describes the largest dynamic route choice model that has been solved to date. Consequently, this research establishes a new benchmark for dynamic route choice modeling. Solving dynamic route choice models on a realistic and large-scale traffic network is no longer an intractable task.

2. This research proposes a link-time-based variational inequality formulation of a DUO route choice problem. Traffic flows propagate over the network in a quasi-continuous manner. FIFO ordering of flows between all O-D pairs are highly maintained to ensure temporally-correct routes and time-continuous flow propagations.

3. The network used in the solution is highly detailed. Links are classified based on geometric characteristics, types of highway facility and types of traffic control. Estimates of delays and travel times of various types of link including turning movements are generated by the proposed model.

4. A key feature of the proposed model is that endogenous and exogenous changes of link capacity are considered. Therefore, adjustments of link capacity are performed to reflect traffic dynamics.

5. Using the proposed model, dynamic network traffic is analyzed which provides a good platform for future ATMS and ATIS applications.

7.2 Future Research

Although this model has successfully demonstrated the capability of solution for a real, large-scale traffic network, several future research needs are identified.

1. This model does not include an integrated traffic signal setting mechanism. The traffic signal timing plan is given as an input and is kept fixed in the solution algorithm. Thus, the dynamic interaction of traffic signals and dynamic traffic flow is thus unable to be analyzed by the current model.

2. Although link capacity is adjusted when queue spillbacks occur, the model does not provide an internal capacity analysis mechanism. Consequently, the capacity used in the solution process might be inappropriate for dynamic traffic conditions and might affect the calculation of intersection delays and link travel times. More precisely, an internal capacity analysis procedure, which is an essential element of a delay model, should be considered in the future.

3. Although realistic traffic engineering-based delay functions are adopted for signalized intersections, the delay functions for freeway-related facilities and unsignalized intersections are still relatively simplistic. Delay functions for freeway-related facilities and unsignalized intersections being tested should be incorporated.

4. The O-D matrices used in this research are based on CATS estimates for 1990. To provide more appropriate estimates of traffic dynamics, time-dependent O-D estimation and forecasting are needed to improve the validity of the model.

5. Unfortunately, link flow and link travel time data are not available for the ADVANCE Network, either in general, or more specifically for the O-D matrix used in this solution. Although the results of this model are internally consistent, they have not been fully validated. These data, as well as route flow data, are urgently needed for the model validation and to advance the state of the art of network modeling for ITS.

6. A more generalized and mathematical-correct dynamic route choice model in consideration of the alternative route choice strategies stated in Section 5.6.1 is still expected to provide a more precise and improved estimates of traffic conditions when unexpected lane-blocking events occur.

7. This model is solved using ten-minute time intervals to perform quasi-continuous dynamic route choice modeling. A sensitivity analysis for different duration of the time interval with respect to the network scale is needed.

8. A sensitivity analysis of using different convergence criterion with respect to the accuracy of the result is needed.

9. A graphical user interface that helps to monitor the dynamic changes in the traffic flow pattern and to visualize the results of the model is worth developing.

10. Parallelization and optimization of the code is needed for the target computing platform to make this code available for application in practice.

References

Akcelik, R. (1988) The Highway Capacity Manual Formula for Signalized Intersections, *ITE Journal*, **58(3)**, 23–27.

Berka, S., Boyce, D.E., Raj, J., Ran, B. and Zhang, Y. (1994) *A Large-Scale Route Choice Model with Realistic Link Delay Functions for Generating Highway Travel Times*, Report to Illinois Department of Transportation, Urban Transportation Center, University of Illinois, Chicago.

Boyce D.E., Kirson, A.M. and Schofer, J.L. (1994) ADVANCE – The Illinois Dynamic Navigation and Route Guidance Demonstration Program, in *Advanced Technology for Road Transport*, I. Catling (ed.), Artech House, Boston, Massachusetts, 247–270.

Boyce D.E., Lee, D.-H., Janson, B.N. and Berka, S. (1995a) Dynamic Route Choice Model of a Large-Scale Urban Traffic Network, Paper presented at the *14th Pacific Regional Science Conference*, Academia Sinica, Taipei, Taiwan.

Boyce D.E., Lee, D.-H., Janson, B.N. and Berka, S. (1995b) Extensions and Application of DYMOD to the ADVANCE Network, Paper presented at the *42nd North American Meeting of the Regional Science International*, Cincinnati; Ohio.

Boyce D.E., Lee, D.-H., Janson, B.N. and Berka, S. (1996) Dynamic User-Optimal Route Choice Modeling of a Large-Scale Traffic Network, Forthcoming in the *ASCE Journal of Transportation Engineering, Special Issue on Advanced Traffic Management and Information Systems*.

Boyce D.E., Ran, B. and LeBlanc, L.J. (1995) Solving an Instantaneous Dynamic User-Optimal Route Choice Model, *Transportation Science*, **29**, 128–142.

Bureau of Public Roads (1964) *Traffic Assignment Manual*, U.S. Department of Commerce, Washington, DC.

Carey, M. (1986) A Constraint Qualification for a Dynamic Traffic Assignment Model, *Transportation Science*, **20**, 55–88.

Carey, M. (1987) Optimal Time-Varying Flows on Congested Networks, *Operations Research*, **35**, 58-69.

Dafermos, S.C. (1980) Traffic Equilibrium and Variational Inequalities, *Transportation Science*, **14**, 42-54.

Dafermos, S.C. and Nagurney, A. (1984) On Some Traffic Equilibrium Theory Paradoxes, *Transportation Research*, **18B**, 101-110.

De Romph, E. (1994) *A Dynamic Traffic Model: Theory and Applications*, Ph.D. Dissertation, Transportation Planning and Traffic Engineering Section, Department of Infrastructure, Delft University of Technology, Delft, Netherlands.

Fisk, C. and Boyce, D.E. (1983) Alternative Variational Inequality Formulations of Network Equilibrium Travel Choice Problem, *Transportation Science*, **17**, 454-463.

Frank, M. and Wolfe, P. (1956) An Algorithm for Quadratic Programming, *Naval Research Logistics Quarterly*, **3**, 95-110.

Friesz, T.L., Bernstein, D., Smith, T.E., Tobin, R.L. and Wie, B.-W. (1993) A Variational Inequality Formulation of the Dynamic Network User Equilibrium Problem, *Operations Research*, **41**, 179-191.

Friesz, T.L., Luque, F.J., Tobin R.L. and Wie, B.-W. (1989) Dynamic Network Traffic Assignment Considered as a Continuous Time Optimal Control Problem, *Operations Research*, **37**, 893-901.

Ho, J.K. (1980) A Successive Linear Optimization Approach to the Dynamic Traffic Assignment Problem, *Transportation Science*, **14**, 295-305.

Ho, J.K. (1990) Solving the Dynamic Traffic Assignment Problem on a Hypercube Multicomputer, *Transportation Research*, **24B**, 443-451.

Janson, B.N. (1991a) Dynamic Traffic Assignment for Urban Networks, *Transportation Research*, **25B**, 143-161.

Janson, B.N. (1991b) A Convergent Algorithm for Dynamic Traffic Assignment, *Transportation Research Record*, **1328**, 69-80.

Janson, B.N. and Robles, J. (1993) Dynamic Traffic Assignment with Arrival Time Costs, *Transportation and Traffic Theory*, Elsevier Science Publishers, Amsterdam, The Netherlands, 127-146.

Janson, B.N. and Robles, J. (1995) A Quasi-Continuous Dynamic Traffic Assignment Model, *Transportation Research Record*, **1493**, 199–206.

Jayakrishnan R., Tsai, W.K. and Chen, A. (1995) A Dynamic Traffic Assignment Model with Traffic-Flow Relationships, *Transportation Research*, **3C**, 51–72.

Kimber, R.M. and Hollis, E.M. (1979) Traffic Queues and Delays at Road Junctions, *TRRL Laboratory Report*, **909**, United Kingdom.

Kyte, M. and Marek, J. (1989) Estimating Capacity and Delay at a Single-Lane Approach All-Way-Stop Controlled Intersection, *Transportation Research Record*, **1225**, 73–82.

Luque, F.J. and Friesz, T.L. (1980) Dynamic Traffic Assignment Considered as a Continuous Time Optimal Control Problem, Paper presented at the *TIMS/ORSA Joint National Meeting*, Washington, DC.

Mahmassani, H.S. and Peeta, S. (1993) Network Performance under System Optimal and User Equilibrium Dynamic Assignments: Implications for ATIS, *Transportation Research Record*, **1408**, 83–93.

Meneguzzer, C., Boyce, D.E., Rouphail, N., and Sen, A. (1990) *Implementation and Evaluation of an Asymmetric Equilibrium Route Choice Model Incorporating Intersection-Related Travel Times*, Report to Illinois Department of Transportation, Urban Transportation Center, University of Illinois, Chicago.

Merchant, D.K. and Nemhauser, G.L. (1978a) A Model and an Algorithm for the Dynamic Traffic Assignment Problems, *Transportation Science*, **12**, 183–199.

Merchant, D.K. and Nemhauser, G.L. (1978b) Optimality Conditions for a Dynamic Traffic Assignment Model, *Transportation Science*, **12**, 200–207.

Nagurney, A. (1993) *Network Economics: A Variational Inequality Approach*, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Patriksson, M. (1994) *The Traffic Assignment Problem: Models and Methods*, VSP, Utrecht, The Netherlands.

Pontryagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V. and Mischenko, E.F. (1962) *The Mathematical Theory of Optimal Processes*, Interscience Publishers, New York.

Ran, B. and Boyce, D.E. (1994) *Dynamic Urban Transportation Network Models*, Springer-

Verlag, New York.

Ran, B. and Boyce, D.E. (1995) Ideal Dynamic User-Optimal Route Choice: A Link-Based Variational Inequality Formulation, *California PATH Working Paper*, UCB-ITS-PWP-**95-7**, Berkeley.

Ran, B., Boyce, D.E. and LeBlanc, L.J. (1993) A New Class of Instantaneous Dynamic User-Optimal Traffic Assignment Models, *Operations Research*, **41**, 192–202.

Ran, B., Hall, R. and Boyce, D.E. (1995) A Link-Based Variational Inequality Model for Dynamic Departure Time/Route Choice, *California PATH Working Paper*, UCB-ITS-PWP-**95-6**, Berkeley.

Ran, B. and Shimazaki, T. (1989a) A General Model and Algorithm for the Dynamic Traffic Assignment Problems, *Transport Policy, Management and Technology Towards 2001, Proceedings of Fifth World Conference on Transport Research*, Yokohama, Japan, 463–477.

Ran, B. and Shimazaki, T. (1989b) Dynamic User Equilibrium Traffic Assignment for Congested Transportation Networks, Paper presented at the *Fifth World Conference on Transport Research*, Yokohama, Japan.

Robles, J. and Janson, B.N. (1995) Dynamic Traffic Modeling of the I–25/HOV Corridor Southeast of Denver, *Transportation Research Record*, **1516**, 48–60.

Smith M.J. (1979) The Existence, Uniqueness and Stability of Traffic Equilibria, *Transportation Research*, **13B**, 295–304.

Smith, M.J. (1993) A New Dynamic Traffic Model and the Existence and Calculation of Dynamic User Equilibria on Congested Capacity-Constrained Road Networks, *Transportation Research*, **27B**, 49–63.

Smulders, S. (1988) Modeling and Filtering of Freeway Traffic Flow, *Transportation and Traffic Theory*, Elsevier Science Publishers, Amsterdam, The Netherlands, 139–158.

Van Aerde, M. and Yager, S. (1988) Modeling Integrated Freeway/Traffic Signal Networks: A Proposed Routing-Based Approach, *Transportation Research*, **22A**, 445–453.

von Stackelberg, H. (1952) *The Theory of the Market Economy*, Oxford University Press, Oxford, United Kingdom.

Wardrop J.G. (1952) Some Theoretical Aspects of Road Traffic Research, *Proceedings of the Institution of Civil Engineers, Part II*, **1**, 325–378.

Webster F.V. (1958) Traffic Signal Settings, *Road Research Laboratory Technical Paper*, **39**, Her Majesty's Stationery Office, London, United Kingdom.

Wie, B.-W. (1989) Dynamic System Optimal Traffic Assignment on Congested Multidestination Networks, Paper presented at the *Fifth World Conference on Transport Research*, Yokohama, Japan.

Wie, B.-W. (1991) Dynamic Analysis of User Optimized Network Flows With Elastic Travel Demand, Paper presented at the *70th Annual Meeting of the Transportation Research Board*, Washington DC.

Wie, B.-W., Friesz, T.L. and Tobin, R.L. (1990) Dynamic User Optimal Traffic Assignment on Congested Multidestination Networks, *Transportation Research*, **24B**, 431–442.

Wie, B.-W., Tobin, R.L. and Friesz, T.L. (1994) The Augmented Lagrangian Method for Solving Dynamic Network Traffic Assignment Models in Discrete Time, *Transportation Science*, **28**, 204–220.

Wie, B.-W., Tobin, R.L., Friesz, T.L. and Bernstein, D. (1995) A Discrete Time, Nested Cost Operator Approach to the Dynamic Network User Equilibrium Problem, *Transportation Science*, **28**, 79–92.

Zhang, Y., Hicks, J. and Boyce, D.E. (1994) Trip Data Fusion in ADVANCE, *ADVANCE Working Paper*, **43**, Urban Transportation Center, University of Illinois, Chicago.