Integration of Dynamic Traffic Control and Assignment

by

Owen Jianwen Chen

B.E. in Management and Systems Science
University of Science and Technology of China, Hefei, China (1991)

M.A. in Economics
Vanderbilt University, Nashville, TN (1993)

M.S. in Transportation
Massachusetts Institute of Technology, Cambridge, MA (1996)

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in
Transportation Systems and Decision Sciences

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1998

Author . . . .
Department of Civil and Environmental Engineering
May 22, 1998

Certified by . . . .
Moshe E. Ben-Akiva
Edmund K. Turner Professor of Civil and Environmental Engineering
Thesis Supervisor

Accepted by . . . .
Joseph M. Sussman
Chairman, Departmental Committee on Graduate Studies
Department of Civil and Environmental Engineering

Integration of Dynamic Traffic Control and Assignment

by

Owen Jianwen Chen

Submitted to the Department of Civil and Environmental Engineering
on May 22, 1998, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Transportation Systems and Decision Sciences

Abstract

This thesis integrates the dynamic traffic control and the dynamic traffic assignment problems. The objective of the combined control-assignment problem is to find a mutually consistent dynamic system-optimal signal setting and dynamic user-optimal traffic flow. Toward this end, we first propose an integrated framework to combine dynamic traffic control and dynamic traffic assignment. Then we develop a game theoretic methodology to model the combined dynamic traffic control-assignment problem as a non-cooperative game between a traffic authority and highway users.

The combined control-assignment problem is first formulated as a one-level Cournot game: the traffic authority and the users choose their strategies simultaneously. The combined control-assignment problem is subsequently formulated as a bi-level Stackelberg game. The traffic authority is the leader; it determines the signal settings in anticipation of the users' reactions. The users are followers who choose their routes after the signal settings have been determined. Finally, the system-optimal control-assignment problem is formulated as a monopoly game. The sole player—the traffic authority—determines both signal settings and traffic flows to achieve a dynamic system optimal solution.

Solution algorithms are developed to solve the combined dynamic traffic control-assignment problem. The developed models and solution algorithms are applied to the Boston Back Bay network. Results from the application demonstrate the potential improvements in traffic performance by integrating control and assignment.

Thesis Supervisor: **Moshe E. Ben-Akiva**
Title: Edmund K. Turner Professor
Department of Civil and Environmental Engineering
Massachusetts Institute of Technology

To my parents

Thesis Committee

Dr. Moshe E. Ben-Akiva (Chairman)

Edmund K. Turner Professor

Department of Civil and Environmental Engineering

Massachusetts Institute of Technology

Dr. Ismail Chabini

Assistant Professor

Department of Civil and Environmental Engineering

Massachusetts Institute of Technology

Dr. Nathan H. Gartner

Professor

Department of Civil and Environmental Engineering

University of Massachusetts–Lowell

Dr. Haris N. Koutsopoulos

Associate Professor

Department of Civil and Environmental Engineering

Carnegie Mellon University

Acknowledgments

I would like to take this opportunity to thank my advisor, Professor Moshe Ben-Akiva, for his invaluable technical advice and guidance throughout my master and doctoral study at MIT. His constant support and encouragement made my stay at MIT the most enjoyable experience in my life.

I would also like to thank other members in my doctoral committee –Professor Ismail Chabini, Professor Haris Koutsopoulos at Carnegie Mellon, Professor Nathan Gartner at University of Massachusetts-Lowell – for their guidance, inspiring advice, and interest in my research work.

My deepest gratitude goes to current and former researchers at the MIT Intelligent Transportation Systems (ITS) Program, especially Dr. Qi Yang, Dr. Michel Bierlaire, Dr. Mithilesh Jha, Mr. Anthony Hotz at MIT Lincoln Lab, Prof. Rabi Mishalani at Ohio State University, Mr. Alan Chachich at Textron, and Mr. Peter Welch at Silicon Graphics. Their technical expertise and help are always available when I need them. I also wish to thank:

The faculty at MIT Center for Transportation Studies, especially Cynthia Barnhart, Nigel Wilson, Joe Sussman, Yosef Sheffi, and Carl Martland.

Professor David Bernstein at Princeton University and Professor Bin Ran at University of Wisconsin–Madison.

Professor Chronis Stamatiadis at University of Massachusetts–Lowell for providing the Boston Back Bay network data.

My fellow students at the MIT ITS Program and the Center for Transportation Studies Kazi Ahmed, David Cuneo, Masroor Hasan, Sridevi Ganugapati, Yiyi He, Winston Guo, Kalidas Ashok, Amalia Polydoropoulou, Constantinos Antoniou, Hariharan Subramanian, Susan Wang, Nagi Rao, Stanley Ouyang, Joan Walker, John Bowman, Jon Bottom and Scott Ramming – for sharing the precious days and nights at 3 Cambridge Center and Room 5-008.

My other fellow students at MIT Center for Transportation Studies Qiang Gao, Yan Dong, Francisco Jauffred, Hong Jin, John Wilson, Jeff Sriver, Jiang Chang, Daniel Rodriguez, Rajesh Shenoi, Chris Caplice, Bill Cowart, Daeki Kim, and Jai-Kue Park – for making my tenure at MIT enjoyable and memorable.

Julie Bernardi and Cheryl Gillespie for their friendship and help.

The UPS Foundation, MIT ITS Program and Bechtel/Parsons Brinckerhoff for their generous financial support through the UPS dissertation fellowship and the Central Artery/Tunnel research project.

Di Wu, for her understanding, patience and unconditional support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The Problem

The development of traffic signal control systems for urban networks has paralleled the development and use of the automobile. The first electric traffic signal in the United States was installed in 1914 (FHWA, 1985) when automobile traffic grew rapidly after World War I. Since then, much effort has been expended to improve the hardware as well as the software to operate traffic signals in a more efficient manner. Today, in typical urban areas, approximately two-thirds of all vehicle miles of travel, and even a higher percent of vehicle-hours of travel, take place on facilities controlled by traffic signals (FHWA, 1985). The quality of traffic control operation has a critical impact on urban traffic system performance.

The techniques available for traffic signal control have progressed considerably since Webster's original study (Webster 1958). Network-wide optimization (Robertson 1969) and traffic responsive control (Hunt et al. 1981; Gartner 1983) are now possible. However, flow inputs for these techniques are usually determined according to traffic counts or predicted flows from traffic assignment models.

On the other hand, traffic assignment models that estimate and predict flow patterns take signal settings as fixed inputs. These two processes, traffic control and traffic assignment, are usually carried out separately, with fixed inputs from one another.

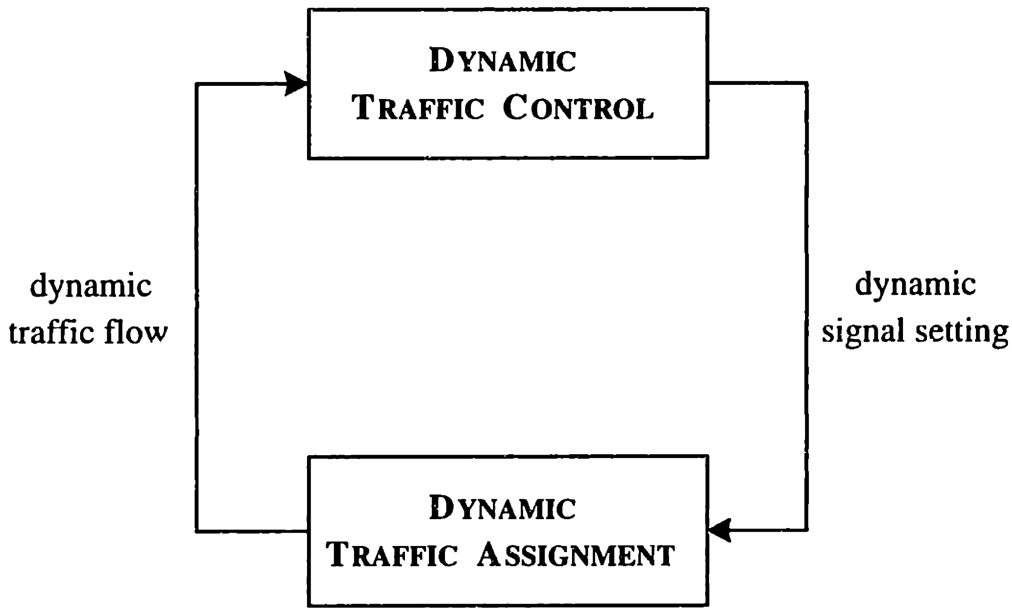

Figure 1-1: Interdependence of traffic control and traffic assignment

In reality, however, drivers will react to changes in signal settings and their reactions will cause changes in traffic flows. The modified traffic flows will subsequently require a change in signal settings and so on. Thus, there exists a strong interdependence between traffic control and traffic assignment, see Figure 1-1. This interdependence is more evident in an envisioned real-time dynamic situation, in which a real-time adaptive traffic control system optimizes traffic signals and drivers actively seek their best routes.

The core of the problem is that the interdependence between control and assignment has been largely ignored in both theory and practice. The direct consequences of such ignorance include: (i) the signal settings given by the control model are non-optimal; and (ii) the traffic flow predictions given by the assignment model are not consistent with users' behavior. The interdependence of control and assignment argues for close coordination of Advanced Traffic Management Systems (ATMS) and Advance Traveler Information Systems (ATIS).

1.2 Objectives of the Thesis

The main objective of this thesis is to find a dynamic system-optimal signal setting and a dynamic user-optimal traffic flow that are mutually consistent. A mutually consistent solution is the state at which a traffic signal control is a *system optimum* for the given traffic flow pattern and the traffic flow pattern is a *user equilibrium* for the given signal control. That is, the signal control and the resulting traffic flow solve both the traffic control and the traffic assignment problems. Such a mutually consistent solution has very important applications in the following two situations:

- The design of a dynamic traffic control system that takes into account drivers' reactions to the signal control.

- The design of a dynamic route guidance system that provides a consistent and unbiased guidance when the signal control is traffic-responsive.

In order to find the mutually consistent signal settings and traffic flows, we need to integrate the dynamic traffic control (DTC) and the dynamic traffic assignment (DTA) problems. Such an integration will help us to design and evaluate traffic control policies and route guidance strategies consistent with user behavior and thus having a sustainable performance over the long term. Specifically, the objectives of this research are:

- To develop a modeling framework to combine signal control and traffic assignment problems in a dynamic situation.

- To formulate a set of analytical models for the combined dynamic traffic control-assignment (DTCA) problem.

- To evaluate existing control policies and route guidance strategies using the developed framework and analytical models.

- To design new control policies that will perform well in the combined control-assignment environment.

- To develop solution algorithms to solve the DTCA problem.

1.3 Thesis Outline

The remainder of the thesis is organized as follows:

Chapter 2 reviews the relevant literature on traffic signal control, traffic assignment and the integrations of the two. Chapter 3 develops a framework to integrate dynamic traffic control (DTC) and dynamic traffic assignment (DTA) systems. Chapter 4 proposes a game theoretic methodology to study the the combined dynamic traffic control–assignment (DTCA) problem. The DTCA problem is modeled as a non-cooperative game between a traffic authority and users. Three sets of models based on Cournot, Stackelberg and Monopoly games are proposed. Chapter 5 develops algorithms to solve the DTCA models. Chapter 6 presents a case study in a Boston local street network and compares results for different models. Chapter 7 concludes and suggests directions for future research.

Chapter 2

Literature Review

This chapter reviews existing studies of traffic signal control, traffic assignment and the integration of control and assignment.

2.1 Traffic Control

The quality of a signal control system is generally measured in terms of *safety* and *efficiency*. The safety objective requires that conflicting traffic movements at an intersection do not have simultaneous right of way. Efficiency in signal control has been attempted through the use of various traffic control strategies.

2.1.1 Signal Control Parameters

A traffic signal control system allocates the right-of-way to conflicting traffic movements. Some fundamental signal control parameters along with their impacts on traffic capacity are described below.

Interval A period of time during which all signal indications remain constant.

Cycle A complete sequence of signal indications.

Phase The portion of a signal cycle allocated to any combination of traffic movements receiving simultaneous right-of-way during one or more intervals.

Cycle Length The total time for a signal to complete one cycle, denoted as C.

Green Time The time within a given phase during which the green indication is shown, denoted as G.

Lost Time The time during which the intersection is not effectively used by any traffic movement. This occurs during the "yellow," "all red" and the beginning of each phase due to start-up delays, denoted as L.

Split The ratio of green time to the cycle length, denoted as $g = G/(C - L)$.

Offset The time difference between the start of the green indication at one intersection and the system reference time.

Saturation Rate The maximum flow rate per lane at which vehicles can pass through a signalized intersection in a stable moving queue. The rate represents the maximum number of vehicles per hour per lane that can pass through a signalized intersection under prevailing traffic and roadway conditions if the green signal were available 100 percent of the time. Denoted as s.

Capacity The maximum number of vehicles per hour per lane that can pass through a signalized intersection under current signal timing, denoted as ca. The capacity of a lane group is the product of the saturation flow rate and the green time split: $ca = s\,g$.

Degree of Saturation The ratio of flow rate (v) to capacity, also known as the *volume-to-capacity ratio*, denoted as ρ. By definition,

$$\rho = \frac{v}{ca} = \frac{v}{s\,g} = \frac{v(C - L)}{s\,G}$$

For any given lane group at a signalized intersection, only three signal indications are seen: green, yellow, and red. The red indication usually includes a short period during which all indications are red, referred to as an "all-red" interval. The yellow and the all-red indications form the clearance and change interval between two green phases. Signal control models differ from one and another in determining the above signal control parameters.

2.1.2 Phasing

Signal phasing can be defined as the way in which the signal cycle is divided to allocate the right-of-way to combinations of traffic movements at an intersection.

A *permissive* turn is a left or right turn at a signalized intersection, made against an opposing or conflicting vehicular or pedestrian flow. By contrast, a *protected* turn is a traffic movement that has the exclusive right-of-way with no opposing or conflicting vehicular or pedestrian flow.

For a typical intersection, there are four approaches with two streets crossing at the intersection. The phasing schemes discussed below apply only to *one street* since it is completely analogous to consider the phasing for the other street. Here, a *street* means the approach under consideration and its opposing approach. Figure 2-1 gives a street layout with two approaches: east bound (EB) and west bound (WB). Each approach has a right turn, a through and a left turn movement. Clearly, the left turns conflict with their opposing through traffic movements. Therefore, the issue of the phasing design is to determine whether or not the left turns need to be protected.

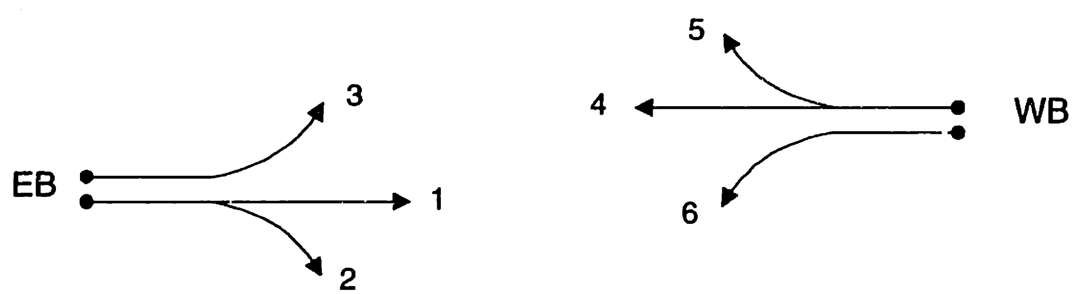

Figure 2-1: A Street with Two Approaches

There many possible phasing schemes for one street, each differing on the left-turn treatment. Common phasing schemes include:

1. **Single**: All movements of EB and WB are allowed at the same time. All left turns are permissive.

2. **Split**: The scheme consists of two phases. In the first phase, all movements of EB are allowed; in the second phase, all movements of WB are allowed. Both left turns are protected.

3. **Leading + Permissive**: A two-phase plan. In the first phase, all movements of EB are allowed and in the second phase, all movements of both EB and WB are allowed. EB's left turn (movement 3 in Figure 2-1) is first protected and then permissive, while WB's left turn (movement 6 in Figure 2-1) is only permissive in phase 2.

4. **Opposing Leading + Permissive**: Similar to scheme 3, but all movements of WB are allowed in the first phase and all movements of WB and EB are allowed in the second phase.

5. **Dual + Permissive**: A two-phase plan. In the first phase, only the left turns of both EB and WB are allowed, while in the second phase, all movements of EB and WB are allowed. Both left turns are protected in the first phase and permissive in the second phase.

6. **Dual + Prohibited**: This is the same as scheme 5, except that the left turns of EB and WB are NOT allowed in the second phase.

7. **Leading + Permissive + Lagging**: A three-phase plan. In the first phase, only movements of EB are allowed. In the second phase, all movements of EB and WB are allowed and in the third phase, only movements of WB are allowed. The left turn of EB is protected in the first phase, permissive in the second phase, and prohibited in the third phase. By contrast, the left turn of WB is prohibited, then permissive, then protected.

8. **Leading + Prohibited + Lagging**: This is a three-phase plan and it is the same as scheme 7, except that in the second phase both left turns are prohibited.

9. **Dual + Leading + Permissive**: This is a three-phase plan. In the first phase, only left turn movements of both EB and WB are allowed. In the second

phase, only movements of EB are allowed. In the third phase, all movements of EB and WB are allowed.

10. **Dual + Opposing Leading + Permissive**: It is a three-phase plan that is similar to scheme 9, except that in the second phase only movements of WB are allowed.

2.1.3 Webster Method

The most commonly used signal timing model is developed by Webster (1958). The Webster model provides a detailed procedure to calculate cycle length and green times. Webster designates a critical lane as the one with the highest ratio of flow to saturation flow. Denote this flow ratio as Y_m for phase m. By definition,

$$Y_m = \max_a \{\frac{v_a}{s_a}\} \tag{2.1}$$

where v_a and s_a are the flow rate and saturation rate for lane a.

The cycle length is calculated as:

$$C = \frac{1.5L + 5}{1 - Y} \tag{2.2}$$

where L is the lost time (in seconds) and Y is the summation of critical flow ratios for all M phases:

$$Y = \sum_{m=1}^{M} Y_m \tag{2.3}$$

The available green time (C–L) is distributed among all phases in proportion to their critical flow ratios:

$$G_m = (C - L)\frac{Y_m}{Y} \tag{2.4}$$

Webster's green time allocation can be characterized as the equalization of the degrees of saturation for the critical movements of all phases. To see this, we denote the flow rate and saturation rate of the critical traffic movement of phase m as v_m and s_m. The degree of saturation for phase m is then given by

$$\rho_m = \frac{v_m}{s_m g_m} = \frac{v_m(C-L)}{s_m G_m} = \frac{v_m(C-L)}{s_m(C-L)\frac{v_m/s_m}{Y}} = Y \tag{2.5}$$

which is the same for every phase m. Therefore, the degrees of saturation for all phases under Webster's control are all equal to the summation of the critical flow ratios.

2.1.4 Three Generations of UTCS Control Techniques

A number of computerized traffic signal control models have been developed since the introduction of microprocessor-based signal controllers in the 1960's. State-of-the-art concepts of traffic-responsive control in urban street networks consist mainly of the three generations of the Urban Traffic Control System (UTCS) (MacGowan and Fullerton 1980). The UTCS project was established in the early 1970's by the Federal Highway Administration (FHWA) to provide for computer-supervised control of 200 signalized intersections in Washington, D.C. The project was directed toward the development and testing of a variety of advanced network control concepts and strategies, and lasted for almost a decade. Its results defined the state-of-the art in the U.S. to the present day (Gartner et al. 1995).

The UTCS research project on control strategies has been divided into three generations of traffic control techniques. These three generations are briefly characterized as follows:

First Generation Control (1-GC)

First-generation control uses pre-stored signal timing plans which are developed off-line based on historical traffic data. The control plan of the system can be selected by time-of-day, direct operator selection, or matching from the existing library a plan that is best suited to recently measured traffic conditions. The plan selection of traffic matching is termed as the traffic-responsive (TRSP) mode, in which the matching criterion is based on a network threshold value which incorporates traffic volumes and occupancies. Timing plans are usually updated once every 15 minutes in the TRSP mode. A transition routine provides a smooth transition between different timing plans. In addition, 1-GC software also includes a critical intersection control (CIC) feature to fine-tune the system at intersections that frequently saturate.

Timing plans for 1-GC are calculated using either an off-line signal optimization method such as TRANSYT (Robertson 1969) or a progression optimization method such as MAXBAND (Little et al. 1981).

Second Generation Control (2-GC)

Second-generation control is an on-line model that computes and implements signal timing plans in real-time based on both surveillance data and predicted traffic data. The software of 2-GC contains an on-line optimization algorithm—SIGOP (Lieberman and Woo 1982), a traffic prediction model, subnetwork configuration models, the critical intersection control (CIC) and a transition model to minimize transition time between two plans. The second generation control package is known as TANSTP: Traffic Adaptive Network Signal Timing Program (MacGowan and Fullerton 1980).

Third Generation Control (3-GC)

The third-generation control is designed to implement and evaluate a fully traffic-responsive, on-line control system. To accomplish this goal, 3-GC is developed to allow the signal timing parameters to change continuously in response to real-time measurements of traffic variables. The time period between timing plan revisions are usually in the magnitude of 3-5 minutes, compared to the frequency of 5-10 minutes of a revision in 2-GC. The cycle length is allowed to vary among the signals as well as at the same signal during the control period. The variable cycle length is accomplished by dividing the control period into an integral number of intersection-specific control intervals, calculated based on the predicted volume/capacity ratios on each approach to the intersection using a Webster-like method.

Table 2.1: Comparison of UTCS Control Techniques

Feature	1-GC	2-GC	3-GC
Update frequency	15 min	5-10 min	3-5 min
Optimization	Off-line	On-line	On-line
Traffic prediction	None	Historically based	Smoothing
Critical intersection control	Split adjustment	Split and offset adjustment	N/A
Cycle length	Fixed within each section	Fixed within groups of intersections	Variable in time & space

A comparison of the major characteristics of the three UTCS generations is shown in Table 2.1.4. Each control generation is designed to provide an increased degree of traffic responsiveness. Prediction of traffic volumes, occupancies and speeds on each link of the street network is critical to the effectiveness of the 2-GC and 3-GC techniques. Results of extensive field testing are counter-intuitive: the fixed-cycle, non-responsive 1-GC performed better than both the 2-GC and the 3-GC (MacGowan and Fullerton 1980). This counter-intuitive outcome is due to the poor performance of the traffic prediction model.

2.1.5 Other Traffic Responsive Controls

Significant progress in the development of advanced traffic responsive control was made during the 1980's in certain parts of the world. In particular, the Sydney Co-ordination Adaptive Traffic System (SCATS) (Sim 1979; Lowrie 1981) was developed in Australia, and the Split, Cycle and Offset Optimization Technique (SCOOT) (Hunt et al. 1981) was developed in England. SCATS is a 1-GC/TRSP variant while SCOOT is a 2.5-GC strategy.

A new family of highly adaptive control methods have been developed over the past decade. The most notable examples of such development include OPAC (Gartner

1983), UTOPIA (Di Taranto and Mauro 1989) and PRODYN (Henry and Farges 1989). Following the OPAC lead, all these strategies adopt a dynamic programming optimization methodology within a rolling horizon framework. UTOPIA was tested in Torino, Italy in 1985-1986. PRODYN was experimentally operated in Toulouse, France and was recently commercialized. OPAC underwent its first experimental test in 1990 (Gartner et al. 1991). The early limited tests showed some great promise but these 3-GC programs have not been fully developed and tested.

2.2 Traffic Assignment

Traffic assignment has been studied intensively in the literature since Wardrop's definition of the traffic equilibrium (Wardrop 1952). Given an O-D demand and the characteristics of the network (including the physical infrastructure and the control policies in effect), a traffic assignment model represents travelers' route choices and predicts the resulting flow pattern over a transportation network.

Transportation planners have traditionally used static traffic assignment to estimate and predict traffic patterns in a transportation network. Methods to solve for static traffic assignment—the traffic is assumed to be in steady-state and flow variables do not change over time—have been well established. Sheffi (1984) provides a comprehensive treatment of the mathematical formulations and the solution algorithms for the static equilibrium problem. Static traffic assignment represents long-run average traffic flow patterns but does not capture the time-varying nature of traffic; it assumes that traffic has reached the steady state. In addition, static traffic assignment assumes a uniform spatial distribution of traffic. Many researchers (Ben-Akiva 1985; Friesz 1985) recognize these unrealistic assumptions of the static traffic assignment and consequently suggest the need for dynamic network analysis.

There has been a significant effort toward the dynamic analysis of traffic networks. The problem has become known as the Dynamic Traffic Assignment or DTA. Unlike static traffic assignment in which the modeling and solution methods are well established and widely accepted, the dynamic traffic assignment problem has been investigated in several different approaches (and many others are under development), with no single generally accepted methodology. Research in this area can be broadly classified in two categories: (1) analytical models that are flow-based, with the mathematical representation of the traffic flow dynamics; and (2) simulation models that are vehicle- or packet-based, with the detailed representation of vehicle or packet movement.

2.2.1 Model Structure

Most existing DTA models, both the analytical and simulation, share the common structure shown in Figure 2-2. A typical DTA model consists of the following components:

- a demand model
- a supply model
- a demand-supply interaction model

The demand model represents users' behavior in the transportation system. The demand is usually given by a set of time-dependent O-D flows and path flows. The set of O-D flows and path flows generated by the demand model often satisfy certain conditions that are consistent with users' behavior (such as deterministic or stochastic route choices).

The supply model represents the network performance. The model includes the network infrastructure (a directed and connected graph consisting of links and nodes), and the performance characteristics of the network (a cost function associated with each link). The supply model generates the network performance for a given demand.

The demand-supply interaction model represents how the supply and demand interact with each other. The interaction produces certain network conditions such as link or path flows and travel times. The network conditions must satisfy both the demand model and the supply model.

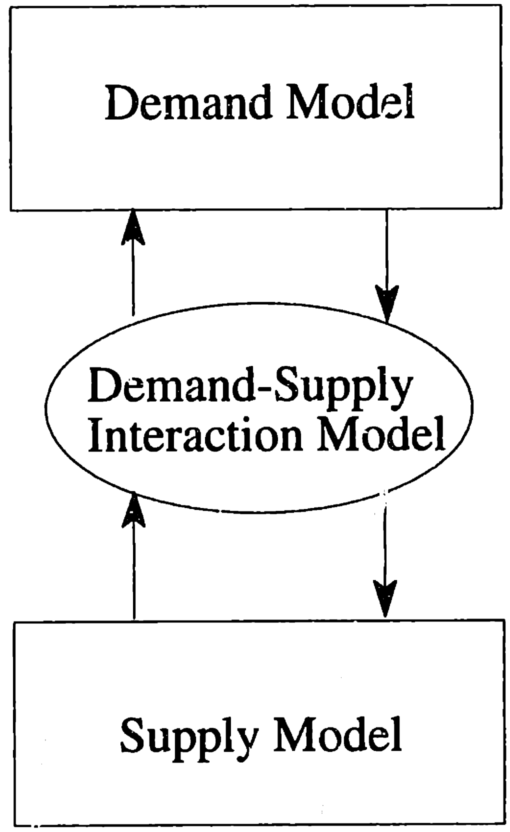

Figure 2-2: A Common Structure of Dynamic Traffic Assignment

The common structure described above will be used to review some representative DTA models in the existing literature.

2.2.2 Flow-based Analytical Models

Flow-based analytical models can be classified into three categories based on the methodology employed: mathematical programming, optimal control and variational inequality models.

Mathematical Programming Models

The first flow-based dynamic traffic assignment model is formulated as a mathematical programming problem by Merchant and Nemhauser (1978a, 1978b). The Merchant-Nemhauser model is a dynamic system-optimal (DSO) route choice problem for a many-to-one network. A time horizon $[0, T]$ is divided into K time subintervals of equal length. The O-D flows are discrete. They are given for each time period and for every origin node. The Merchant-Nemhauser model is given by the following mathematical programming problem:

(P1)

$$\min \sum_k \sum_a X_{a,k}(x_{a,k}) \tag{2.6}$$

subject to

$$x_{a,k} - x_{a,k-1} = u_{a,k} - w_a(x_{a,k-1}) \tag{2.7}$$

$$\sum_{a \in A(i)} u_{a,k} - \sum_{a \in B(i)} w_a(x_{a,k-1}) = p_{i,k} \qquad \forall a \in N - \{q\} \tag{2.8}$$

$$\sum_{a \in A(q)} u_{a,k} - \sum_{a \in B(q)} w_a(x_{a,k-1}) + s_k = 0 \tag{2.9}$$

$$u_{a,k} \geq 0 \tag{2.10}$$

$$x_{a,k} \geq 0 \tag{2.11}$$

where

N	=	set of nodes in the network;
$A(i)$	=	set of incoming links to node i;
$B(i)$	=	set of outgoing links from node i;
q	=	single destination node;
$x_{a,k}$	=	number of vehicles in link a at interval k;
$u_{a,k}$	=	number of incoming vehicles to link a at interval k;
$w_a(x_a)$	=	exit link function of link a;
$p_{i,k}$	=	number of vehicles entering the network at node i at interval k;
s_k	=	number of vehicles leaving the network from destination q at interval k;
$X_{a,k}(x_{a,k})$	=	cost function for link a at interval k.

In terms of the DTA common structure in Figure 2-2, the demand model of users' behavior in the above Merchant-Nemhauser model is substituted by a system-optimal route choice assumption. Since the total system cost is minimized, the resulting network condition is system-optimal.

The supply model is formulated as a set of network flow constraints, which are constructed with link variables. The first group of constraints are state equations expressing the conservation of vehicles on link a. The second group of constraints are balance equations of flows at non-destination nodes. The last equation is the balance of flows at the destination node q.

The demand-supply interaction model is not explicitly represented, but the demand and supply models interact with each other through the link volumes and link cost functions.

There are two major limitations in the Merchant-Nemhauser model. First of all, users are assumed to follow non-realistic system-optimal paths and consequently the demand model of user behavior is missing. Second, the Merchant-Nemhauser model considers only a many-to-one network that has a single destination.

Carey (1987) reformulates the Merchant-Nemhauser problem to a well-behaved convex nonlinear program. In order to overcome the non-convexity issue of the Merchant-Nemhauser formulation, Carey includes an exit flow variable $v_{a,k}$ as a decision variable for each link a in the formulation. The exit flows on links are nonnegative and are bounded from above by the exit link functions $w_a(x_a)$. This modification leads to a convex nonlinear programming model with substantial advantages. For instance, good solution algorithms are available since a piecewise linearized version is a standard linear program. This program automatically satisfies the ordered set property. Moreover, both necessary and sufficient conditions exist to characterize the optimal solution.

Still, Carey's model suffers the limitations inherited from the Merchant-Nemhauser model, including the system-optimal solution and the lack of an explicit demand-supply interaction mechanism. Although this reformulation resolves the issue of non-convexity, it brings an undesirable property known as the FIFO (first-in-first-out) discipline violation. Subsequently, Carey (1992) presents four classes of constraints that ensure the FIFO discipline of multi-commodity flows. Each of the four classes of constraints results in a non-convex constraint set and nonlinear integer program.

Carey suggested an empirical solution as follows: First, solve the dynamic assignment problem without constraints to ensure the FIFO discipline. Then analyze the degree of overtaking or FIFO violations and introduce FIFO constraints if necessary.

Janson (1991) presents a dynamic traffic assignment model as a discrete time optimization problem, in which the solution algorithm consists of the once-through incremental assignment. In a later work (Janson and Robles, 1993), the model is extended to include departure time choice, in which departure time is affected by the arrival time cost. The time horizon is divided into discrete sub-periods with predetermined departure rates for each sub-period. The conditions for user equilibrium are a generalization of the static user equilibrium conditions. The solution algorithm is designed to produce assignments that approximate such conditions.

Ben-Akiva *et al.* (Ben-Akiva, Cyna and de Palma 1984 and Ben-Akiva, de Palma Kanaroglou 1986) develop a framework for analyzing the dynamic traffic assignment problem with stochastic route and departure time choices. This approach incorporates a dynamic driver behavior model and a dynamic network performance model. The route and departure time decisions are depicted using a probabilistic path and departure time choice model, in which a driver's choice depends on perceived travel times for alternative routes and departure times. Moreover, a schedule delay function penalizes late or early arrivals at the destination. The major advantage of this approach is that it presents a general structure for demand-supply interactions in dynamic traffic models. The work is later extended by Vythoulkas (1990) to a general network.

Optimal Control Models

The Merchant-Nemhauser model of dynamic system-optimal traffic assignment is further extended to a multiple origin-destination (O-D) network by Friesz *et al.* (1989), Wie (1989) and Ran and Shimazaki (1989a) using optimal control theory. The Merchant-Nemhauser model is reformulated as a continuous time optimal control problem. The optimality conditions are derived using the Pontryagin's Minimum Principle.

Friesz *et al.* (1989) propose a dynamic user-optimal (DUO) route choice model by considering the equilibration of instantaneous route costs. The instantaneous DUO route choice problem is to determine vehicle flows at each time instant on each link when each driver chooses the route that minimizes instantaneous travel time from origin to destination at the instant of departure time.

In terms of the structure in Figure 2-2, the demand model of Friesz *et al.* (1989) is formulated as minimizing the summation over all links, of the integral of link cost function:

$$\min z(x) = \sum_a \int_0^T f(x_a(t))dt$$

Variable t denotes a time instant. The cost function, denoted by $f(x_a(t))$, represents the instantaneous cost on link a when it contains x users at time t. The function is considered to be a continuous, increasing and convex function. The significant improvement of Friesz's model over the Merchant-Nemhauser model is that users' route choice behavior is modeled.

A generalized DUO route choice model over a multiple origin-destination network is presented by Wie *et al.* (1990). The Wardropian equilibrium principle is generalized in the dynamic case to represent the temporal evolution of traffic flow pattern:

> *If, at each time instant, for each origin-destination pair, the instantaneous expected unit travel costs for all the paths that are being used are identical and equal to the minimum instantaneous expected unit path cost, the corresponding time-varying flow pattern is said to be user optimized (Wie* et al., *1990).*

The expected unit travel cost in Wie *et al.* (1990) is assumed to be a function of traffic volume. That is, the expected unit travel cost of link a is defined as: $c_a[x_a(t)]$, where $x_a(t)$ is the number of vehicles traveling on link a at time t. The dynamic user-optimal traffic assignment model is formulated as the following continuous time optimal control problem:

(P2)

$$\min J = \sum_a \int_0^T \int_0^{x_a(t)} \xi_a c_a(w) dw dt \tag{2.12}$$

subject to

$$\dot{x}_a^n(t) = u_a^n(t) - \xi_a x_a^n(t) \tag{2.13}$$

$$S_{kn}(t) + \sum_{a \in B(k)} \xi_a x_a^n(t) - \sum_{a \in A(k)} u_a^n(t) = 0 \tag{2.14}$$

$$u_a^n(t) \geq 0 \tag{2.15}$$

$$x_a^n(0) = x_a^{n,0} \geq 0 \tag{2.16}$$

where

$A(k)$	=	set of incoming links to node k;
$B(k)$	=	set of outgoing links from node k;
$x_a^n(t)$	=	number of vehicles in link a with destination n at time t;
$u_a^n(t)$	=	inflow rate on link a with destination n at time t;
$c_a[x_a(t)]$	=	expected unit travel cost of link a;
ξ_a	=	a positive constant;
$x_a^{n,0}$	=	number of vehicles on link a with destination n at time *0*;
$S_{kn}(t)$	=	O-D demand with origin k and destination n at time t.

The performance index J is a scalar function defined as the summation over all arcs of appropriate definite integral of $\xi_a c_a[x_a(t)]$. The performance index J is purely instrumental. It has no economic interpretation and is analogous to the objective function of Beckmann's equivalent optimization problem (Beckmann *et al.*, 1956) for a static user equilibrium traffic assignment.

In the above optimal control models (Friesz *et al.*, 1989 and Wie *et al.*, 1990), link inflow, $u_a(t)$, is the only control variable, while link volume (i.e., number of vehicles on a link), $x_a(t)$, is a state variable. The exit flow from a link is modeled as a function of the number of vehicles on that link, $x_a(t)$. This function is known as the exit flow function. In Wie *et al.* (1990), the exit flow function is defined as a scalar function of the link volume: $\xi_a x_a(t)$, where ξ_a is a positive constant. When such an exit flow function is non-linear, the model becomes very complicated and it is very difficult to establish an optimization model for the dynamic user-optimal (DUO) route choice problem on a multiple origin multiple destination network.

Ran and Shimazaki (1989b) present a DUO route choice model in which the exit flow is redefined as a control variable, rather than a derived state variable. Although such a model overcomes the difficulty of the exit flow function, it introduces a new problem as well: overtaking. That is, vehicle movements within a link may violate the first-in, first-out (FIFO) discipline. A situation may occur in which vehicle A enters a link *after* vehicle B but vehicle A exits *before* vehicle B on that link.

Subsequently, Ran *et al.* (1993) improves the Ran and Shimazaki model by adding a flow propagation constraint that prevents the overtaking. The users' behavior model is refined to correspond to the following instantaneous dynamic user-optimal condition:

> *The dynamic traffic flow over the network is in a dynamic user optimal state if for each O-D pair at each decision node at each instant of time, the travel times for all routes that are being used equal the minimum instantaneous route travel time.*

This definition differs from the one given in Friesz *et al.* (1989) and Wie *et al.* (1990) in that users can have route choices not only at the origin node but also at the intermediate nodes, and the optimal condition holds for the intermediate nodes as well.

The link instantaneous travel time in Ran *et al.* (1993) is defined as $c_a[x_a(t), u_a(t), v_a(t)]$, a function of number of vehicles $x_a(t)$, entry flow $u_a(t)$, and exit flow $v_a(t)$ on link a at time t. This instantaneous link travel time is the time that a vehicle would spend on the link if traffic conditions remain unchanged while traversing the link. It is further assumed that the link travel time, $c_a(t)$ is the sum of two components:

$$c_a(t) = g_{1a}[x_a(t), u_a(t)] + g_{2a}[x_a(t), v_a(t)] \tag{2.17}$$

The equivalent optimal control model of the instantaneous dynamic user-optimal route choice problem is formulated as follows:

(P3)

$$\min_{u,v,x,e,E} \int_0^T \sum_a \left\{ \int_0^{u_a(t)} g_{1a}[x_a(t), w]\, dw + \int_0^{v_a(t)} g_{2a}[x_a(t), w]\, dw \right\} dt \qquad (2.18)$$

s.t.

Relationship between state and control variables:

$$\frac{dx_{ap}^{rs}(t)}{dt} = u_{ap}^{rs}(t) - v_{ap}^{rs}(t), \qquad \forall a, r, s \qquad (2.19)$$

$$\frac{dE_p^{rs}(t)}{dt} = e_p^{rs}(t), \qquad \forall p, r, s \qquad (2.20)$$

Flow conservation constraints:

$$\sum_{a \in A(r)} \sum_p u_{ap}^{rs}(t) = D^{rs}(t), \qquad \forall r, s \qquad (2.21)$$

$$\sum_{a \in B(j)} v_{ap}^{rs}(t) = \sum_{a \in A(j)} u_{ap}^{rs}(t), \qquad \forall j, p, r, s; j \neq r, s \qquad (2.22)$$

$$\sum_{a \in B(s)} \sum_p v_{ap}^{rs}(t) = e^{rs}(t), \qquad \forall r, s \qquad (2.23)$$

Flow propagation constraints:

$$x_{ap}^{rs}(t) = \sum_{b \in \bar{p}_j} \{x_{bp}^{rs}[t + \bar{\tau}_a(t)] - x_{bp}^{rs}(t)\} + E_p^{rs}[t + \bar{\tau}_a(t)] - E_p^{rs}(t),$$

$$\forall a \in B(j); j \neq r; p, r, s \qquad (2.24)$$

Definitional constraints:

$$u_a(t) = \sum_{r,s} u_{ap}^{rs}(t), \qquad \forall a \tag{2.25}$$

$$v_a(t) = \sum_{r,s} v_{ap}^{rs}(t), \qquad \forall a \tag{2.26}$$

$$x_a(t) = \sum_{r,s} x_{ap}^{rs}(t), \qquad \forall a \tag{2.27}$$

Boundary conditions:

$$E_p^{rs}(0) = 0, \qquad \forall p, r, s \tag{2.28}$$

$$x_{ap}^{rs}(0) = 0, \qquad \forall a, p, r, s \tag{2.29}$$

Non-negativity conditions:

$$x_{ap}^{rs}(t), u_{ap}^{rs}(t), v_{ap}^{rs}(t) \geq 0, \qquad \forall a, p, r, s \tag{2.30}$$

$$e_p^{rs}(t), E_p^{rs}(t) \geq 0, \qquad \forall p, r, s \tag{2.31}$$

where,

$\mathcal{P}^{rs}$ = set of all paths from node r to node s;

$\mathcal{A}(i)$ = set of links whose tail node is i;

$\mathcal{B}(i)$ = set of links whose head node is i;

$x_a(t)$ = number of vehicles on link a at time t;

$u_a(t)$ = entry flow on link a at time t;

$v_a(t)$ = exit flow on link a at time t;

$E_p^{rs}(t)$ = cumulative vehicles arrived at s from r along path p by time t;

$e_p^{rs}(t)$ = number of vehicles arrived at s from r along path p at time t;

$D^{rs}(t)$ = demand from origin r to destination s at time t;

The objective function is a generalization of the well-known static user-optimal model and is similar to the performance index function of J in Wie *et al.* (1990). The control variables are $u_{ap}^{rs}(t), v_{ap}^{rs}(t), u_{ap}^{rs}(t)$ and $e_{ap}^{rs}(t)$. The state variables are $x_{ap}^{rs}(t)$ and $E_p^{rs}(t)$.

With respect to the supply modeling, Ran *et al.* (1993) add a new constraint—flow propagation constraint for each link. The flow propagation constraint states that vehicles on a link using a given route at any time must result in either added vehicles on downstream links (links on a sub-route following the arc) at a posterior instant of time, or added exiting vehicles at the destination. Thus, flows on the links are forced to remain on the arc for an amount of time consistent with the link travel time. This is an improvement in flow propagation modeling with respect to the previous models, in which an exit flow function is used.

Variational Inequality Models

Variational inequality (VI) problems are being used extensively in formulating economic and engineering problems in new ways that lead to tractable solution algorithms. Applications in transportation engineering are relatively recent, beginning in 1980 (Dafermos, 1980).

Several researchers have shown Beckmann's mathematical programming formulation of Wardrop's principle of static traffic assignment to be a special case of a more general variational inequality formulation (see the review by Friesz, 1985). The variational inequality approach to the traffic assignment problem has two major advantages over the mathematical programming and the optimal control approach. First, a variational inequality formulation is more general and it can handle the asymmetric traffic assignment problem that does not have a valid optimization formulation. Second, there exist solution algorithms to solve a VI formulation directly.

Friesz *et al.* (1993) present a variational inequality formulation of dynamic user equilibrium with simultaneous route and departure time choices. The model is a generalization of Ben-Akiva's dynamic peak-period model (Ben-Akiva *et al.*, 1984, 1986a and 1986b). As in Ben-Akiva's model, users of a transportation network make their rational choices of route and departure time in order to minimize each individual's travel disutilities. Friesz *et al.* show that the simultaneous route and departure time equilibrium problem is equivalent to the following variational inequality problem:

$$\sum_{p\in P}\int_0^T C_p(t,h^*)[h_p(t)-h_p^*(t)]dt \geq 0, \quad \forall h \tag{2.32}$$

where $h_(t)$ is the flow on path p at time t and $C_p(t,h)$ is the path travel cost on path p at time t.

The supply modeling of Friesz *et al.* (1993) is characterized by a linear link delay function, $D_a[x_a(t)]$ as follows:

$$D_a[x_a(t)] = \alpha x_a(t) + \beta \tag{2.33}$$

where $\alpha, \beta > 0$.

In fact, the above linear delay function is similar to the deterministic queuing delay function in Ben-Akiva's model (Ben-Akiva *et al.*, 1984, 1986a and 1986b).

Given the link traversal time, $D_a[x_a(t)]$, the link exit time function can be determined by:

$$\tau_{a_1}^p(t) = t + D_{a_1}[x_{a_1}(t)] \qquad \forall p \in P \tag{2.34}$$

$$\tau_{a_i}^p(t) = \tau_{a_i-1}^p(t) + D_{a_i}[x_{a_i}(\tau_{a_i-1}^p(t))] \qquad \forall i \text{ on path } p, p \in P \tag{2.35}$$

The main advantage of the linear delay function of D is that the link exit time function τ is strictly increasing under the linear D and hence the inverse function τ^{-1} exists. The existence of the inverse exit time function ensures that the first-in, first-out (FIFO) queuing discipline is maintained. This is because the strict monotonicity of the link delay function guarantees that overtaking cannot occur on any link. See Ben-Akiva and de Palma (1986) for a discussion of overtaking.

Jauffred and Bernstein (1996) extend Friesz's simultaneous route-departure model to the case in which vehicles are integer-valued, instead of real-valued, flows.

Ran and Boyce (1996) present a set of variational inequality models for both deterministic and stochastic dynamic user-optimal route choice problems. The dynamic network flow constraints are the same as those in the MP formulation of (**P3**) (Equations 2.19-2.31). The path-based variational inequality formulation is given by the following form:

$$\int_0^T \sum_{r,s} \sum_p \eta_p^{rs^*}(t) \left[h_p^{rs}(t) - h_p^{rs^*}(t)\right] dt \geq 0, \qquad \forall\, h_p^{rs} \tag{2.36}$$

where $\eta_p^{rs}(t)$ is the travel disutility for path p from origin r to destination s at time t. For the instantaneous dynamic user-optimal (DUO) route choice problem, $\eta_p^{rs}(t)$ is defined as the instantaneous path travel time, which is the sum of the instantaneous link travel times for all links on that path. For the realized DUO (or ideal DUO as in Ran and Boyce, 1996) route choice problem, $\eta_p^{rs}(t)$ is defined as the travel time actually experienced over path p from origin r to destination s by vehicles departing origin r at time t. Assume that path p consists of nodes $(r,1,2,...,i,...,s)$. Then a recursive formula for path travel time $\eta_p^{rs}(t)$ is given by:

$$\eta_p^{ri}(t) = \eta_p^{r(i-1)}(t) + \tau_a[t + \eta_p^{r(i-1)}(t)], \qquad \forall p, r, i;\ i = 1, 2, ..., s; \tag{2.37}$$

The stochastic dynamic user-optimal route choice problem is modeled in Ran and Boyce (1996) using the logit function. The perceived path travel time is assumed to be of the form:

$$\Omega_p^{rs}(t) = \eta_p^{rs}(t) - \frac{1}{\theta}\epsilon_p^{rs}(t) \qquad \forall r, s, p \tag{2.38}$$

where ϵ is a nonnegative parameter that scales the perceived travel time and $\epsilon_p^{rs}(t)$ is a random error term. Furthermore, $\epsilon_p^{rs}(t)$ is assumed to be identically and independently distributed (IID) Gumbel variates for each time instant with location parameter equal to zero and scale parameter θ (see Ben-Akiva and Lerman, 1985). The probability of choosing path p at time t for a traveler departing at origin r heading to destination s is thus given by the following multinomial logit choice function:

$$P_p^{rs}(t) = \frac{exp[-\theta\eta_p^{rs}(t)]}{\sum_q exp[-\theta\eta_q^{rs}(t)]} \quad \forall r,s,p \tag{2.39}$$

The stochastic dynamic user-optimal (SDUO) route choice conditions are then defined as follows:

$$h_p^{rs}(t) = D^{rs}(t)P_p^{rs}(t) \quad \forall r,s,p \tag{2.40}$$

where $D^{rs}(t)$ is the dynamic O-D demand and $h_p^{rs}(t)$ is the path flow. In general the mean actual path travel time $\eta_p^{rs}(t)$ is increasing with path flow, $h_p^{rs}(t)$. That is,

$$\frac{\partial\eta_p^{rs}(t)}{\partial h_p^{rs}(t)} > 0 \quad \forall r,s,p \tag{2.41}$$

For each path p and each O-D pair r-s, define an artificial cost term as

$$G_p^{rs}(t) = [h_p^{rs}(t) - D^{rs}(t)P_p^{rs}(t)]\frac{\partial\eta_p^{rs}(t)}{\partial h_p^{rs}(t)} \quad \forall r,s,p \tag{2.42}$$

Clearly, $G_p^{rs}(t) = 0$ because of the SDUO route choice conditions (Equation 2.40). The variation inequality formulation of the SDUO route choice problem is thus given by

$$\int_0^T \sum_{r,s}\sum_p G_p^{rs}(t)\left\{h_p^{rs}(t) - h_p^{rs*}(t)\right\} dt \geq 0 \quad \forall h \tag{2.43}$$

where superscript * denotes the optimal solution.

Recent Analytical Models

Recently, Chabini and He (1998a, 1998b) and He (1997) proposed a conceptual framework and a flow-based analytical DTA model. The framework consists of four major components: users' behavior model, dynamic network loading model, link performance model and path generation module. Three types of users are considered: (1)

following fixed route, (2) choosing the route with the minimum perceived travel time, and (3) choosing the route with the minimum actual travel time. The route choice behavior of the multiple-class users is formulated as a variational inequality problem. Similar to those by Friesz et al. and Wu et al., the dynamic network loading component is formulated as a system of equations representing the link dynamics, flow conservation, flow propagation and boundary constraints. Compared to previous models, Chabini and He's model has two major advantages:

1. the model is valid even if the FIFO condition is not satisfied; and
2. the model does not depend on any particular link performance function.

Friesz et al. (1993) assumed a linear queuing delay function while Wu et al. (1995) and Xu et al. (1996) relaxed to a non-linear volume-delay function. In Chabini and He (1998a, 1998b), the link performance function is a modular component that is user-defined.

By using efficient data structures and special computational methods, Chabini and He (1998a, 1998b) developed and implemented efficient solution algorithms for the DTA model. Through two computational examples, Chabini and He (1998a, 1998b) have demonstrated that the analytical DTA model can be solved faster than real time.

2.2.3 Vehicle-based Simulation Models

While analytical DTA models represent vehicles as continuous flows, the simulation-based DTA models represent vehicles as discrete individual vehicles or groups of vehicles.

CONTRAM

The first vehicle-based DTA model used in real network studies is CONTRAM (Leonard *et al.*, 1989), a dynamic extension of an early static CONTRAM model (Leonard *et al.*, 1978). The model handles traffic in groups called packets. Each packet is processed as a single undivided entity and is assumed to experience the same traffic conditions. Each packet is loaded onto the network by an incremental assignment. The movement of packets through the network is determined by statistical and queuing theory relations. The conditions a packet will encounter on its route depend on the order it is assigned to the network. In an iterative process, packets are removed and re-assigned to the network in consecutive iterations, with the expectation that the iterative process may converge to a dynamic equilibrium.

INTEGRATION

Van Aerde *et al.* (1992) develop the INTEGRATION program to simulate both freeway and arterial networks. INTEGRATION implements traffic assignment by using dynamic probit loading on preset fixed routes. Various types of responsive and fixed-time traffic signals and ramp meters are incorporated in the model as well. The control system, however, functions independently from the traffic assignment. The program can simulate platoon dispersion/coordination on arterial networks and perform shock wave queuing analysis on freeways.

Smith's Model

Smith (1993) proposes a DTA model by using different methods to represent the demand and the supply. The demand model is formulated as an equivalent variational inequality, an equivalent fixed point problem or an equivalent minimization problem. With the fixed point formulation and the continuous route cost function of the route inflows, Smith proved that a dynamic route equilibrium exists.

Smith represents the supply with a macroscopic traffic model that is similar to the CONTRAM model. Vehicles are grouped together as a packet. A packet is regarded as a set of vehicles that have certain common characteristics. For example, they all follow the same route, arrive at the same node at a given time, and experience the same delay. The packet volume is usually not an integer. The model is capacity-constrained, meaning that a node restricts its output due to its exit capacity. In order to exactly fulfill the node exit capacity, it is often necessary to split a packet so that part of the packet can proceed while the remaining part stays at the node. Split packets follow the same route. The packet splitting implies that parts of the same original packet may reach the destination at different times. Priorities are introduced to place packets of traffic queuing at the same node at the same time in a proper order. As time proceeds, packet priorities are adjusted so as to ensure that, as traffic traverses the network, there is no overtaking. Time is discrete and the length of a time interval is defined as the free-flow travel time needed to traverse each unit link. The packets are moved through the network time-epoch by time-epoch and the average route costs are computed for each route and departure time.

DYNASMART

Mahmassani *et al.* (1992) present an assignment-simulation model called DYNASMART. Given a network configuration and a time-dependent O-D demand, DYNASMART simulates traffic patterns and evaluates system performance. Traffic flow is represented in a hybrid fashion in which vehicles can be tracked individually or in macro-particles. Traffic flow moves according to the relationship between speed and density on a link in a macroscopic traffic. Vehicle paths are modeled explicitly as the outcome of individual path selections on each node in the network. Vehicles are classified as follows: Class-1 vehicles follow prescribed system-optimal paths. Class-2 vehicles follow user-optimal paths. Class-3 vehicles follow a boundedly rational switching rule in responding to descriptive information on prevailing conditions. Class-4 vehicles are unguided and follow externally specified paths. The signal control in DYNASMART includes pretimed and actuated controls. For the ATIS application, a rolling horizon approach is implemented. Using the method of successive average (MSA), the model is solved for UE (user equilibrium), SO (system optimum), or a combination of the two, with the hope that the solution will converge.

DYNA

Ben-Akiva *et al.* (1994a) and Cascetta *et al.* (1994) develop a real-time dynamic traffic assignment model as part of the European DRIVE-II project DYNA. DYNA is designed to provide a rolling horizon prediction of network conditions in terms of speed, flows, queue lengths and travel times. Vehicles are modeled in a continuous packet approach. Vehicles within a packet are uniformly distributed between the head and the tail of the packet. The model traces the movement of each packet along its followed path by keeping track of its head and tail. The model allows for multiple user classes with different types of available information and/or travel behavior rules. Discrete choice models are adopted to represent more realistic path and departure-time choices. The model also allows for en route diversion of paths based on provided real-time information.

MITTNS

Ben-Akiva *et al.* (1994b) develop a mesoscopic simulation-based model—MITTNS (MIT Traffic Network Simulator). MITTNS consists of the following major modules:

- Real-time Origin-Destination Estimation and Prediction Module;
- Driver Behavior Module;
- Traffic Routing Module;
- Traffic Simulation Module;
- Traffic Control Interface Module;

The MITTNS model's core is its mesoscopic time-based traffic simulation module, which follows a packet approach as found in other simulation-based models. However, MITTNS adds an important feature to control running time. That is, vehicle packets can be aggregated into *traffic cells*. While packets are indivisible and cannot be split, traffic cells can be merged or split dynamically with the smallest traffic cell being one packet. Links are further divided into segments that can be generated dynamically. Each segment consists of a moving part and a queuing part. Signals, closed lanes, and incidents are represented by their effects on the aggregate capacity at the downstream end of the segment. Each segment has an output capacity and an acceptance capacity. The dynamic segment representation facilitates the modeling of queue formation, dissipation and spill-back. Furthermore, traffic streams are adopted to model lanes. Velocity of a traffic packet on a segment is assumed to be a function of the traffic density ahead of the packet on the moving part of the segment. The simulation proceeds in two phases: (i) the *update phase* in which the values of the model's traffic variables such as speeds, densities, and queue lengths are updated and (ii) the *advance phase* in which the traffic packets are advanced to their new positions.

DynaMIT

Recently, Ben-Akiva *et al.* (1997) developed another DTA system at MIT—DynaMIT. DynaMIT is designed as a real-time Dynamic Traffic Assignment system that provides traffic predictions and travel guidance. Travel guidance refers to information provided to a trip-maker in an attempt to facilitate his/her decisions on departure time, travel mode and route. Clearly, departure time and (to some extent) travel mode recommendations are only effective prior to trip departure, whereas route guidance information may be useful both before and during a trip.

DynaMIT is designed to operate in real time, accept real-time surveillance data, and estimate and predict time-dependent O-D flows. The system incorporates different driver classes and their behavior, provides self-calibration capabilities, estimates current network conditions, predicts future traffic conditions, interfaces with the traffic control system, and generates route guidance consistent with the predicted traffic conditions. It is designed to make departure time, mode and route recommendations for a variety of information systems and information dissemination strategies.

As shown in Figure 2.2.3, DynaMIT consists of two major components:

- Current State Estimation;
- State Prediction and Guidance Generation.

State Estimation provides estimates of the current state of the network in terms of O-D flows, link flows, queues and densities, given historical and surveillance data. The estimation is based on explicit and consistent simulation of the interaction between network supply and traveler demand. The state prediction and guidance generation

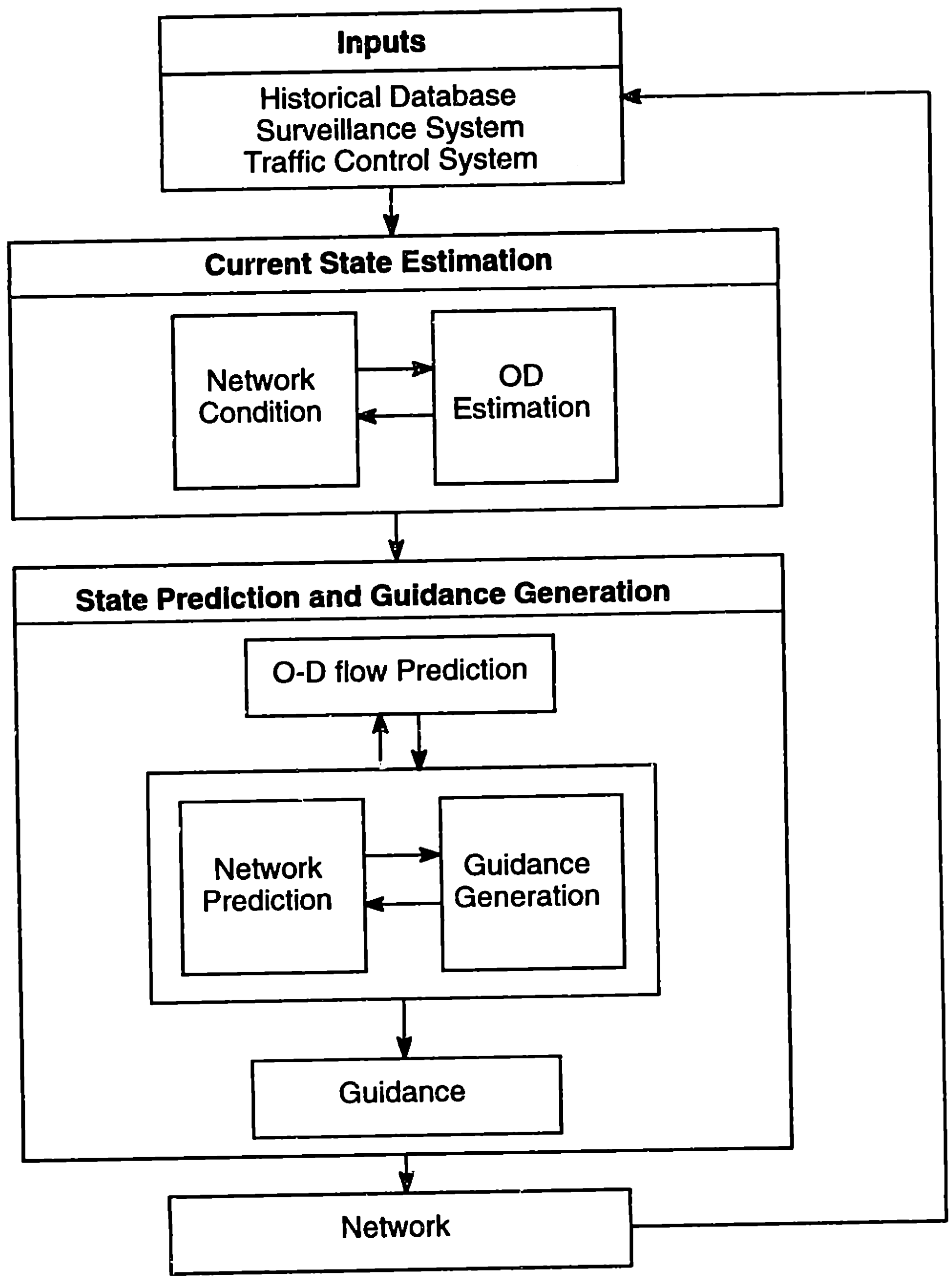

Figure 2-3: DynaMIT System Structure

component takes the state estimate as input and generates anticipatory guidance. The system enforces consistency between the travel times on which the guidance is based and the travel times that result from travelers' reactions to the guidance.

These two components (state estimation and prediction-based guidance generation) utilize various models and algorithms to attain their objectives. Some of these models and algorithms are shared by the two components. DynaMIT features the following models:

- a dynamic O-D matrix estimation and prediction model which estimates and predicts time-dependent O-D flows by applying Kalman filter methods to real-time data from traffic sensors and identifying deviations from historical traffic patterns;

- a disaggregate-level demand simulator which explicitly models the pre-trip departure time, mode and path choice behavior and the *en route* switching behavior of a number of user classes;

- a mesoscopic-level traffic simulator which moves individual vehicles or vehicle packets according to macroscopic traffic relations; and

- a guidance generation model which generates both anticipatory descriptive and prescriptive guidance that is consistent and unbiased.

2.3 Integration of Traffic Control and Assignment

Until very recently, little attention has been paid to the significance of the relationship between signal control and traffic assignment; these have been studied as two mutually independent traffic engineering activities. All the traffic signal control models reviewed in Section 2.1 optimize signals for the current flow pattern, which is assumed not to change as a result of the altered signal settings. Similarly, all the traffic assignment models reviewed in Section 2.2 assume that signal settings are exogenously given. The interaction between control and assignment was first addressed in the literature in the mid-1970s. Since then, researchers have begun to realize the importance of integrating traffic control and assignment.

Two approaches have addressed the interaction between control and assignment: (i) the iterative procedure that solves control and assignment problems iteratively, and (ii) the global optimization approach that aims at the global optimality of the control policies taking users' route adjustment into account.

2.3.1 Iterative Procedure

Allsop (1974) is among the first to address the interaction between traffic control and traffic assignment. He argues that the two processes could not be considered independently and suggests that traffic equilibrium theory should embrace the signal-setting process. He expresses the link cost functions with explicit reference to the green times:

$$\tau_a(u_a, g_a) \tag{2.44}$$

where τ_a is the link travel cost which is a function of both link flow u_a and link green time g_a. Through this link cost function, Allsop suggests that the traffic signals can, in principle, be used to influence the link and path travel costs and thus to influence the traffic flow pattern.

Maher and Akcelik (1975) observe that an area traffic control policy could have re-distributional effects on traffic flows. They point out one potential danger of the advanced traffic-responsive signal control if it is not carefully designed: the continual optimization of signal plans can lead to a gradual deterioration of the network performance. They conclude that the benefits of an advanced traffic control scheme could not be made properly without consideration of the re-distributional effects. In later studies (Akcelik and Maher, 1977 and Maher and Akcelik, 1977), they demonstrate through simulation experiments that potential savings in travel time result from the use of route control techniques in an urban street network where the dominant control mechanism is traffic signal control.

Gartner (1976 and 1977) also emphasizes the need to incorporate the traffic equilibrium analysis into the area traffic control design. He suggests a procedure to incorporate route choice into the traffic signal optimization process. The suggested procedure is the inclusion of traffic flows as decision variables in the objective function of the usual traffic signal timing problem. Thus, rather than assuming that the set of traffic flows is known and fixed, Gartner suggests that the traffic flow variables be determined simultaneously with the traffic signal control variables.

Allsop and Charlesworth (1977) conduct a number of computer tests, in which the iterative control assignment procedure is applied using TRANSYT on a small six-junction network. The network has a fixed set of static trip demands. They start the experiments by first fixing the signal settings and solving the traffic assignment. Then they call the traffic signal optimization program, TRANSYT, to optimize the signal settings according the new set of link flows. The procedure is repeated with the new signal settings, resulting in a new set of link flows and so on. These tests show that for the combined control-assignment problem, multiple mutually consistent points might exist where the flows are in user equilibrium and signals are optimal for those flows. Although these solutions have quite different flow and green time patterns, the total travel times in the network are virtually equal. They also find that the final solution for the iterative procedure depends on the initial signal settings.

Smith (1979) shows through a simple two-link example that the interaction of Webster's equal-saturation control policy and route choice does not solve the optimum control problem. To develop an alternative to Webster's policy, Smith (1979) proposes a local control policy, P_0. For a simple intersection with two separately controlled approaches, Smith's P_0 policy is defined as follows:

$$s_1 \tau_1 = s_2 \tau_2 \tag{2.45}$$

where s_1 and τ_1 are the saturation flow rate and link delay for approach 1, and s_2 and τ_2 are the corresponding variables for approach 2. Later Smith (1981a) generalizes the above policy to multiple phases by defining the *cost of the clique* [1] as

$$C^i = s_i \tau_i \tag{2.46}$$

Policy P_0 is to equalize the cost of the clique, C^i for every clique i of the intersection. Equivalently, policy P_0 implies that *less costly cliques receive no green time.* Smith (1981b) shows that policy P_0 possessed the capacity-maximization property. This property is achieved by attracting traffic to higher capacity links (larger s_i), as the control policy favors links with high saturation flow rates by allocating proportionally more green time to those links. This is consistent with the observation by Heydecker (1983) that signal control policies should discourage drivers from using roads with high marginal costs. Similar to Webster's equal-saturation method, Smith's P_0 is a local control policy and therefore does not necessarily minimize the total cost in a network.

Gartner *et al.* (1980) propose two iterative procedures for modeling the combined route choice and traffic control problem. Both procedures iteratively solve the traffic assignment problem and the traffic signal setting problem. However, two different programs are used to solve the signal setting problem. The first procedure, assuming non-interconnected signals, uses Webster's formulas (Webster, 1958) to calculate cycle length and green splits. The second procedure adopts a signal network optimization program, MITROP (Gartner *et al.*, 1975), to solve the signal setting problems.

Dickson (1981) shows via a simple example that the iterative control-assignment procedure does not necessarily result in mutually consistent green time/equilibrium flow combinations that minimize the total travel time in the network. He argues that an iterative approach might lead to a decline rather than an improvement in network performance. He suggests that signal settings should be adjusted in such a way to take into account the resulting changes in the equilibrium flows. However, he gives no indication in the paper of how this could be done.

[1] A *clique* of approaches is a set of approaches containing no conflicting pair. It is equivalent to the concept of a *phase*.

Cantarella *et al.* (1991) propose another iterative procedure for the solution of the combined traffic signal control and traffic assignment problem. The proposed procedure is similar to those mentioned above, except that the traffic control problem is solved in two stages. The first stage determines the green splits and their sequences based on capacity maximization. The second stage determines the signal coordination based on total delay minimization. They test the procedure on some small networks and find that it does not have a satisfactory convergence property.

Al-Malik and Gartner (1995) and later Gartner and Al-Malik (1996 and 1997) investigate the Wardropian traffic equilibrium under Webster control. They modify Beckmann's traffic assignment formulation by incorporating the green times as the decision variables. These green times are constrained to satisfy Webster's equal-saturation formulas. The resulting problem is essentially a symmetric traffic assignment with two conflicting approaches of traffic. The mathematical formulation of the problem is similar to Dafermos' extended traffic assignment model with link interaction (Dafermos 1971). Signal coordination is also incorporated in an extended model (Gartner and Al-Malik, 1997). A *progression adjustment procedure* that calculates offsets by using TRANSYT and modifies link delay through a *progression adjustment factor* is added to the link-interaction assignment model.

The iterative control and assignment studies focus on the consistency of control policies, that is, the long-term effects that signal control policies have on the patterns of traffic flow. Essentially, the control and assignment are formulated as two interrelated problems in the same level and these two problems are solved iteratively until the consistency between the two solutions is met. However, there is no guarantee that the iterative procedure will yield a converged solution. Furthermore, even if the procedure does converge, it is likely that a local optimal solution rather than a global one is reached.

2.3.2 Global Optimization

To overcome the iterative procedure's shortcomings of non-convergence and local optimality, a few researchers have adopted a different approach—global optimization.

The global optimization models aim for global optimality of the control policies taking into account travelers' route adjustment.

A group of researchers at the Massachusetts Institute of Technology (Gershwin *et al.* 1978, Tan *et al.* 1979, and Gartner *et al.* 1980) were among the first to formulate mathematically the combined traffic control and assignment problem. They formulated a network traffic control problem that incorporates individual drivers' route choice behavior as a *Hybrid Optimization Program* (HOP). They recognized that the objectives of the traffic authority and the drivers are different. The HOP formulation is similar to a bi-level programming problem, but the lower level program is represented by equilibrium traffic flows that satisfy individual drivers' route choice behavior. A Generalized System Optimization Problem (GSOP) is also introduced when the traffic authority could use both control parameters and traffic flows as decision variables to optimize some system-wide cost. Two extended equilibrium principles are derived for HOP and GSOP, respectively, and these two principles are very similar to the equal travel time of user equilibrium and the equal marginal cost of the system optimum. An iterative optimization-assignment algorithm with an augmented Lagrangian method is used to solve the HOP problem in some simple networks.

Sheffi and Powell (1983) analyze the optimal signal timing problem in a network. They attribute the non-convergence of the iterative procedure reported in the literature to the difference between a user equilibrium flow pattern and a system optimizing flow that a traffic control authority seeks when setting signal timing. They formulate the optimal signal setting problem as a mathematical program in which the traffic flow is constrained to be user equilibrium. A heuristic is proposed to solve the network signal optimization formulation.

Smith and Ghali (1990) and Smith and Van Vuren (1993) also combine static traffic assignment with responsive signal control. They investigate theoretically various signal control policies, including Webster, P_0, delay minimization, and other gradient-like policies. These control policies differ from one another only in the definitions of the "pressure" for each phase. They extend Smith's (1981) finding on policy P_0 to any control policy: *less pressurized phase receives no green time.* This optimality condition for the signal control problem is equivalent to the user equilibrium condition for traffic assignment: *more costly path has no traffic flow.* Smith and Van Vuren (1993) also show that the link cost function plays an important role in determining a control policy's theoretical properties and thus its practical performance results.

Recently, Yang and Yagar (1994, 1995) formulated the combined control and assignment problem as a bi-level program in which the upper level is the optimal traffic control problem and the lower level is the static user equilibrium assignment problem. In order to accommodate the signal control impacts, Yang and Yagar modify the traffic assignment model with queues. Whenever a link traffic flow is greater than the link exit capacity, the portion of the link flow over capacity is put into the queue. It is not intuitive that the queuing delay is an endogenous variable rather than a function of flow and capacity. The queuing delay is determined by the network equilibrium. A sensitivity analysis method is employed to obtain the derivatives of the equilibrium link flows and equilibrium queuing delays with respect to the signal settings. The derivative information is then used in a gradient descent algorithm to solve the bi-level formulation.

With two exceptions, all the above-mentioned studies consider only the static traffic assignment model. One exception is Gartner and Stamatiadis (1996) who propose a framework to integrate dynamic traffic assignment with real-time control. The second exception is Chen and Hsueh Chen and Hsueh (1997) who formulate a mathematical program for a dynamic traffic-responsive signal control scheme based on the Webster delay function along with a deterministic user-optimal route choice behavior. The model is only solved on a simple network.

In this thesis, we will use game theory as a methodology to integrate traffic control and assignment in a dynamic setting. Game theory provides a framework for modeling a decision-making process in which multiple players are involved. Game theory first appears in transportation problems in the form of the so-called Wardropian equilibrium (Wardrop, 1952) of route choice, which is similar to a *Nash equilibrium* of a n-player, non-cooperative game (Nash 1950). Fisk (1984) formulates the signal optimization problem as a Stackelberg game in which a regulating agency minimizes a network performance function and travelers choose static user-optimal routes. He proposes two procedures to solve a Stackelberg-type problem. One is the inner-outer iterative scheme and the other is a penalty-function approach.

2.4 Summary

Traffic engineering studies on both signal control and traffic assignment have a long and rich history. Research and development of traffic control strategy has evolved from simple local, pre-timed control to advanced network, traffic-responsive control. In the traffic assignment area, development of *dynamic* models has become prominent in this field. Both the flow-based analytical DTA models and the vehicle-based simulation models have their own advantages. Recent development shows that an analytical DTA model may be usually easier to solve while a simulation model can offer much richer users' behavior and more realistic system performance.

By contrast, there are only limited studies in the literature on the integration of the control and assignment problems. The majority of the existing studies focus on the iterative procedure. Most of the studies only model the static traffic situations. Clearly, there is a need to provide a comprehensive study on the integration of dynamic traffic control and dynamic traffic assignment.

Chapter 3

A Framework for Integrating Dynamic Traffic Control and Dynamic Traffic Assignment

This chapter develops an integrated framework to combine the dynamic traffic signal control and the dynamic traffic assignment problems. The framework combines two separate traffic engineering efforts on control and assignment in order to find an optimal signal control plan, taking into account individual users' travel decisions.

3.1 An Integrated Framework

As reviewed in Chapter 2, traffic control and traffic assignment are two very important and interdependent traffic engineering systems. Each has been developed independently. An integrated framework is proposed to bridge the gap between these two systems.

Figure 3-1 illustrates an integrated Dynamic Traffic Management System (DTMS). The Traffic Management Center (TMC) consists of two main components: Dynamic Traffic Control (DTC) and Dynamic Traffic Assignment (DTA). The TMC collects real-time information on traffic conditions through a surveillance system and uses the measurements for both the DTC and DTA. On the one hand, the DTC of the TMC

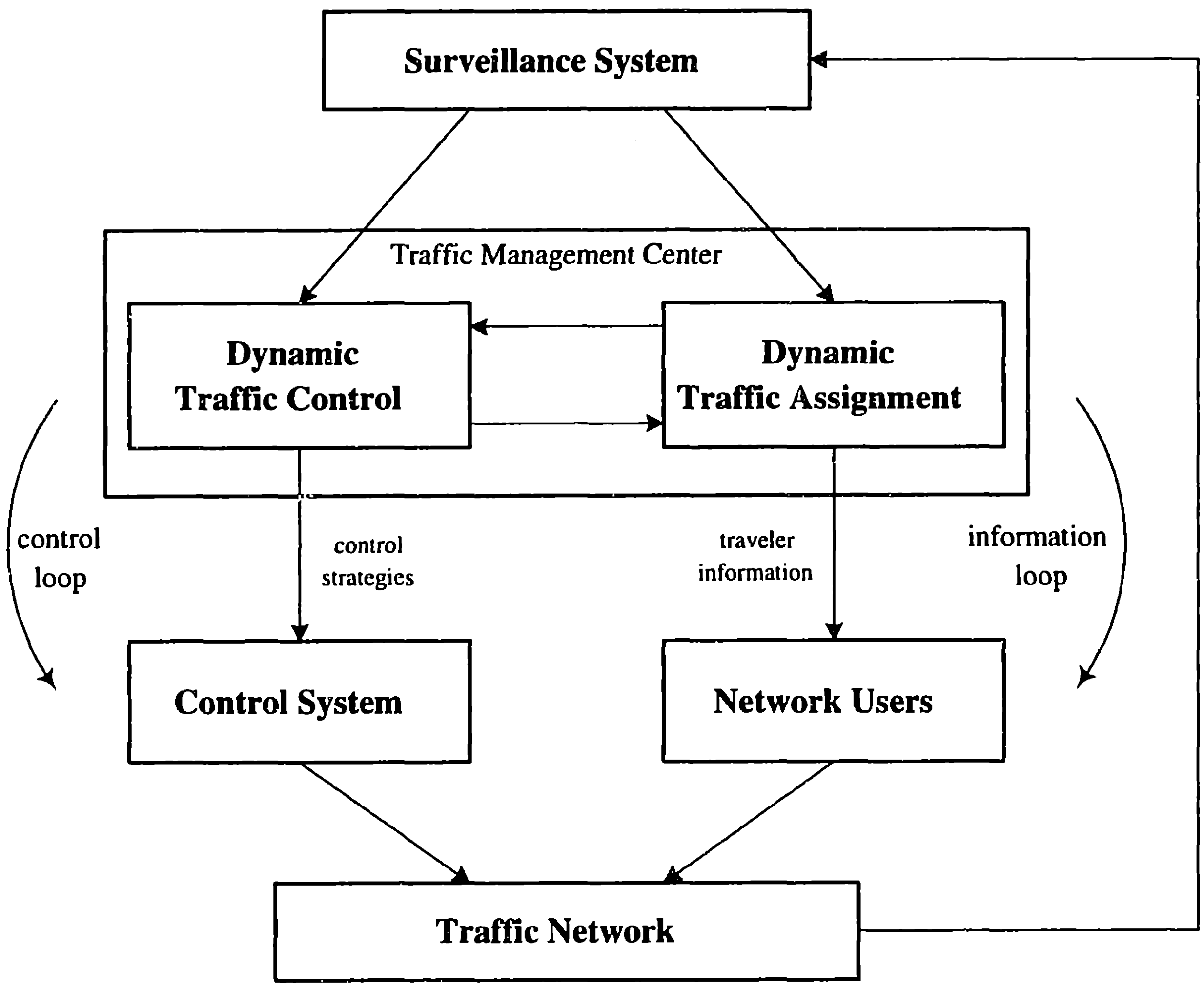

Figure 3-1: An Integrated Dynamic Traffic Management System (DTMS)

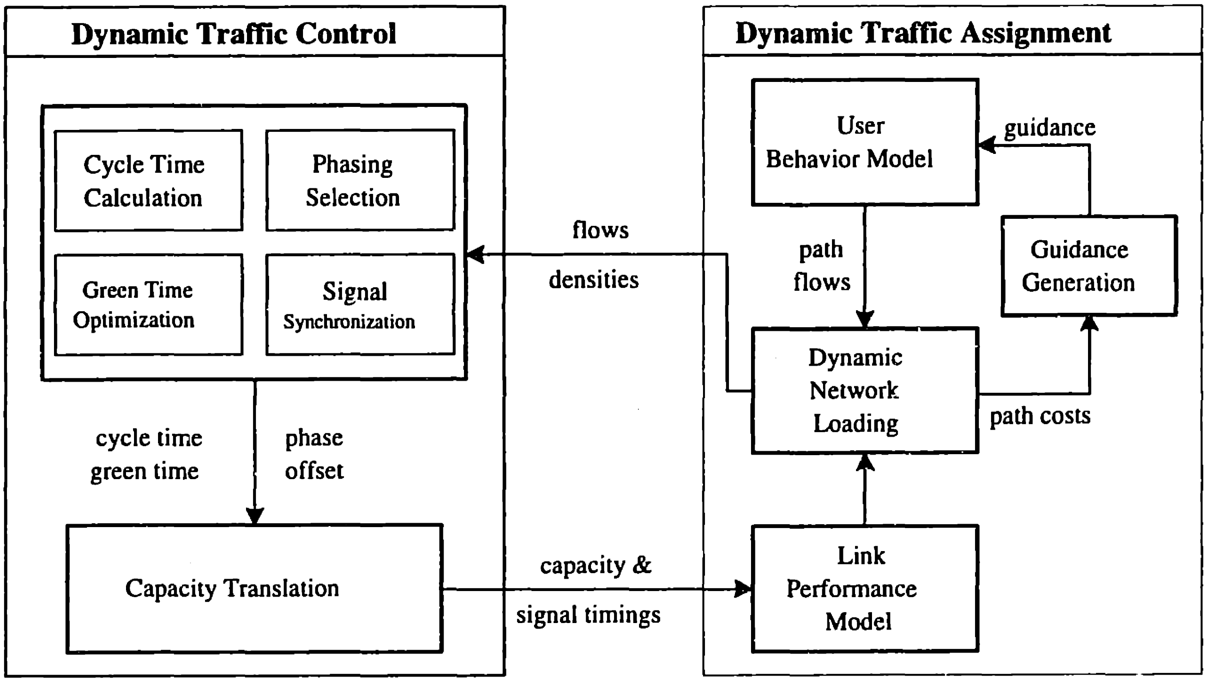

Figure 3-2: An Integrated Framework for Dynamic Traffic Control and Assignment

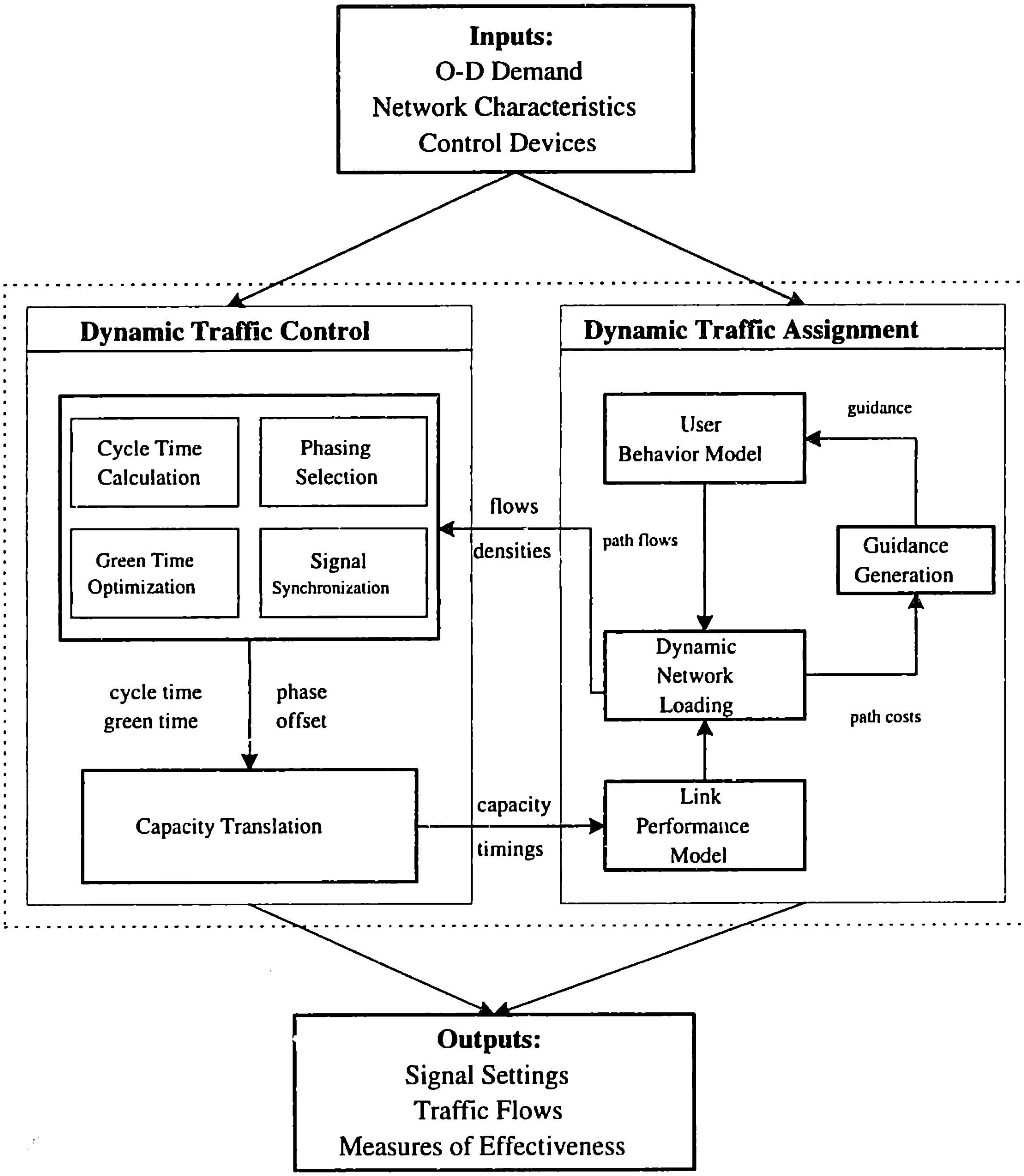

Figure 3-3: An Off-line System

optimizes real-time control strategies that are transmitted via the control system to the traffic network and affect the network supply. On the other hand, the DTA of the TMC models the users' behavior and generates dynamic guidance information that is provided to network users and affects the network demand.

Our focus is on integrating the two main components of the TMC: DTC and DTA. Figure 3-2 illustrates an integrated framework for combining DTC and DTA. The interaction between DTC and DTA components can be viewed as the demand-supply interaction. DTA takes the network supply such as capacity and signal timing as inputs, provides travel guidance to users and predicts network usage for the given supply. By contrast, DTC takes the network demand such as link and path flows as inputs, optimizes the control system and thus determines the network supply. The resulting network condition is therefore an equilibrium of the demand–supply interaction. Chapter 4 will provide a methodology and some analytical models to capture such a demand–supply interaction.

The integrated framework is designed for both off-line planning and on-line operating purposes. Figure 3-3 shows how the framework is implemented in an off-line system. Inputs to the off-line system include O-D demand, network characteristics, and control devices. Once a mutually consistent solution is found between the DTC and the DTA, the system produces outputs of system-optimal signal settings, user-optimal traffic flows and measures of effectiveness (MOEs). We focus on the off-line system in this thesis.

For an on-line system shown in Figure 3-1, the O-D estimation and prediction must be built into the system. In addition, some implementation techniques such as *rolling horizon* need to be incorporated into the on-line system.

3.2 Dynamic Traffic Control System

A dynamic traffic control (DTC) system is an extension of a traditional traffic control system to the dynamic situation. As shown in Figure 3-4, DTC contains the following four signal optimization models:

- Phase Selection
- Cycle Time Calculation
- Green Time Optimization
- Signal Synchronization

As reviewed in Chapter 2, a traffic control system must optimize a set of parameters. The above four signal optimization models are associated with the determination of the corresponding signal timing parameters: phase, cycle, green, and offset, respectively. Figure 3-4 also shows that each of the four models do not exist in isolation; rather, they interacts with each other. The detailed functionalities of the four models are described next.

3.2.1 Phase Selection

Phasing is a technique to reduce conflicts between traffic movements at signalized intersections. The selection of an appropriate signal phasing is usually the first critical process of a traffic signal design. Signal phasing can be defined as the way in which the signal cycle is divided to allocate the right-of-way to combinations of traffic movements at an intersection.

A *permissive* turn is a left or right turn at a signalized intersection, made against an opposing or conflicting vehicular or pedestrian flow. By contrast, a *protected* turn is a traffic movement that has the exclusive right-of-way with no opposing or conflicting vehicular or pedestrian flow. Generally, a left turn conflicts with an opposing through traffic movement. Therefore, the issue of the phasing design is to determine whether or not left turns need to be protected.

Section 2.1.2 reviews possible phasing schemes for a typical intersection. The method for phase selection is heuristics-oriented. After identifying possible phasing schemes, the system performance for each is evaluated for the given traffic flow and other signal timing parameters (cycle and green time). The phasing plan with the best system performance is then selected.

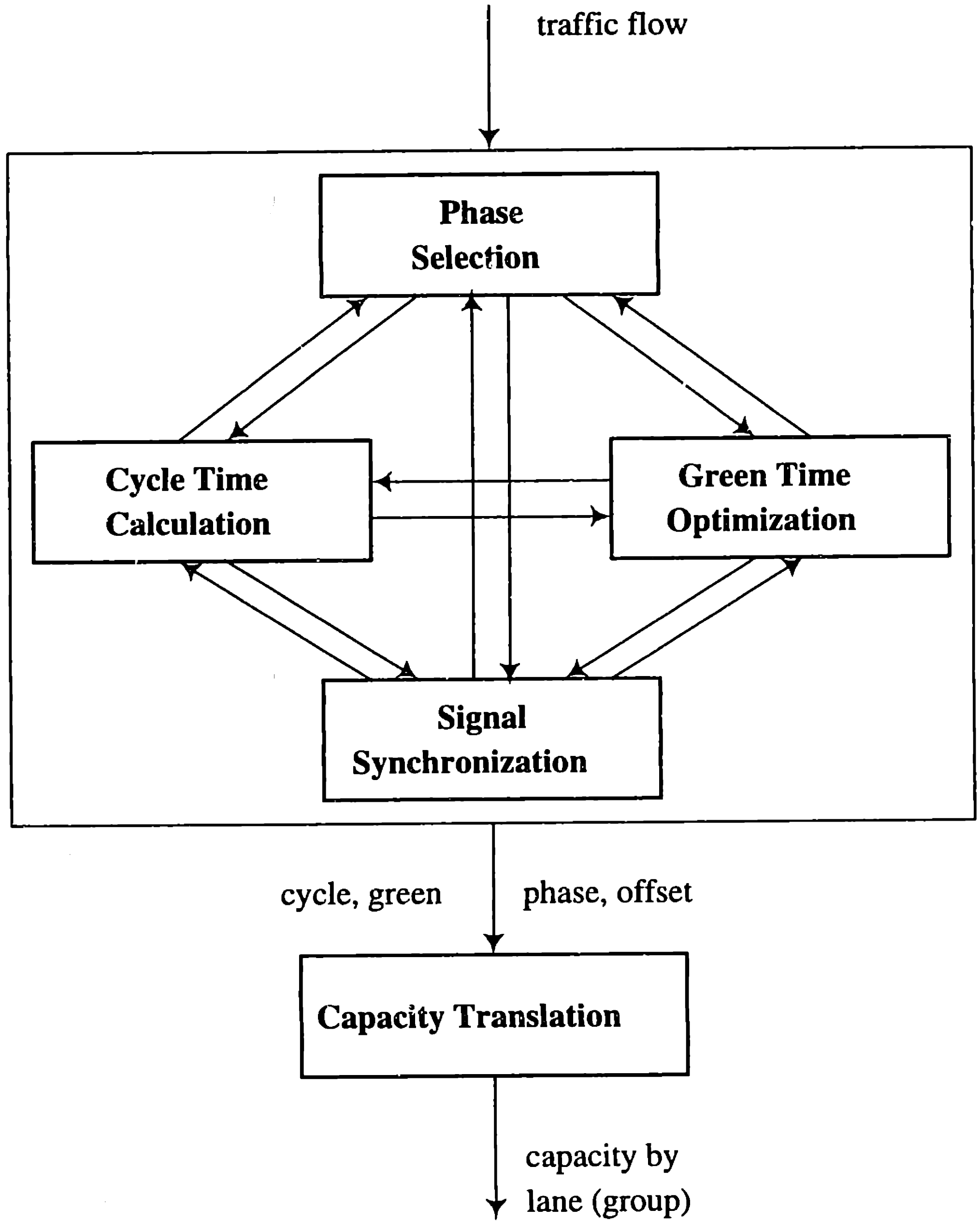

Figure 3-4: Dynamic Traffic Control (DTC) System

3.2.2 Cycle Length Calculation

The indications of a traffic signal in an intersection are cycled in green, yellow and red. A cycle length is defined as the time during which a complete sequence of signal phases occur.

As reviewed in Section 2.1.3, the most prominent method for cycle length calculation is the Webster formula (Equation 2.2). We will use the Webster formula to calculate the cycle length at a given isolated intersection i.

3.2.3 Green Time Optimization

Green time is the most important parameter in a traffic control system because it has the most significant influence on traffic capacity. In fact, a link capacity is proportional to the green time allocated to that link[1]. As reviewed in Chapter 2, there are several methods for green time optimization, including the popular Webster formula and recent Smith's P_o policy. In this thesis, however, the green time optimization problem is formulated as a mathematical programming problem as follows:

Minimize Total Delay
subject to
phasing constraint
cycle length constraint
minimum green time constraint

Detailed mathematical formulation will be given in Chapter 4. Since the objective is to minimize the total delay (or total travel time) in the network, such a green time policy is also known as the *delaymin* policy.

3.2.4 Signal Synchronization

Signal synchronization is a very important element of signal design for both an arterial street and an urban traffic network.

[1]Remember that a link capacity is defined as the product of the saturation flow rate and the green time split for that link.

The signal design for an arterial street is concerned with controlling signals along the arterial street in order to give primary consideration to the provision of progressive traffic flow along the arterial. In contrast to an isolated intersection signal design, the arterial street signal design gives rise to a unique arterial traffic characteristic that vehicles traveling along the arterial are released in *platoons* from an upstream signal, and move in platoons to the next downstream signal. Thus it becomes desirable to establish a time relationship between the beginning of the arterial green at one intersection and the beginning of the arterial green at the next intersection, so that vehicles traveling in platoons can receive a green indication without stopping on any signalized intersection along the arterial. This continuous progression of traffic flow along an arterial will significantly reduce the delays on the arterial.

The technique for establishing a signal synchronization along an arterial relies on the time-space diagram. Using the time-space diagram, the through-band[2] is maximized by adjusting the offsets of the signals along the arterial. Since the offset relationship must be obtained over time, it is necessary to have a common cycle or some multiple of a common cycle throughout the arterial.

3.2.5 Capacity Translation

The capacity of a lane is defined as the maximal flow that can pass the lane *given* existing traffic, roadway and signal setting conditions. Capacity basically depends on the saturation flow rate and the effective green time available to the particular lane. Therefore, the capacity of lane a of intersection i at time t, $c_a(t)$, is defined as:

$$c_a(t) = s_a \frac{G_a(t)}{C_i(t)} \tag{3.1}$$

where,

[2] A ***through-band*** is defined as the space between a pair of parallel speed lines which delineates a progressive movement on a time-space diagram. The ***slope*** of the through-band represents the progressive speed of traffic moving along the arterial and the ***bandwidth***—the width of the through-band—measures the period of time available for traffic to progress within the band.

$$s_a \quad = \quad \text{saturation flow rate for lane } a \text{ (veh/hr/lane);}$$
$$G_a(t) \quad = \quad \text{effective green time for lane } a \text{ during time } t;$$
$$C_i(t) \quad = \quad \text{cycle length for intersection } i \text{ during time } t.$$

The cycle length, $C_i(t)$, is determined by Equation 2.2 of Section 3.2.2 and the effective green time, $G_a(t)$, is given by the green time optimization model that is described in Section 3.2.3 and will be discussed in detail in Chapter 4. The saturation flow rate for lane a, s_a, is the flow that could be accommodated by lane a assuming the green phase was available 100 percent of the time. Factors that affect a lane's saturation flow rate include lane width, approach grade, parking activity, left or right turn, etc.

3.3 Dynamic Traffic Assignment System

A dynamic traffic assignment (DTA) system provides time-dependent travel guidance and predicts future traffic conditions given time-dependent O-D demand, signal settings and network capacity. Figure 3-5 gives the structure of the DTA system that contains the following four components:

- User Behavior Model
- Guidance Generation
- Link Performance Model
- Dynamic Network Loading Model

The user behavior model takes dynamic O-D demand and travel guidance as inputs and assigns the time-dependent O-D departure rate to a set of available paths for each O-D pair according to users' route choice behavior. This results in a set of time-dependent path flows, which are inputs to the dynamic network loading (DNL) model. The DNL model loads the path flows to the network such that a set of network flow constraints are satisfied. The outputs of the DNL are time-dependent link flows and link costs which are inputs to the guidance generation model. The guidance

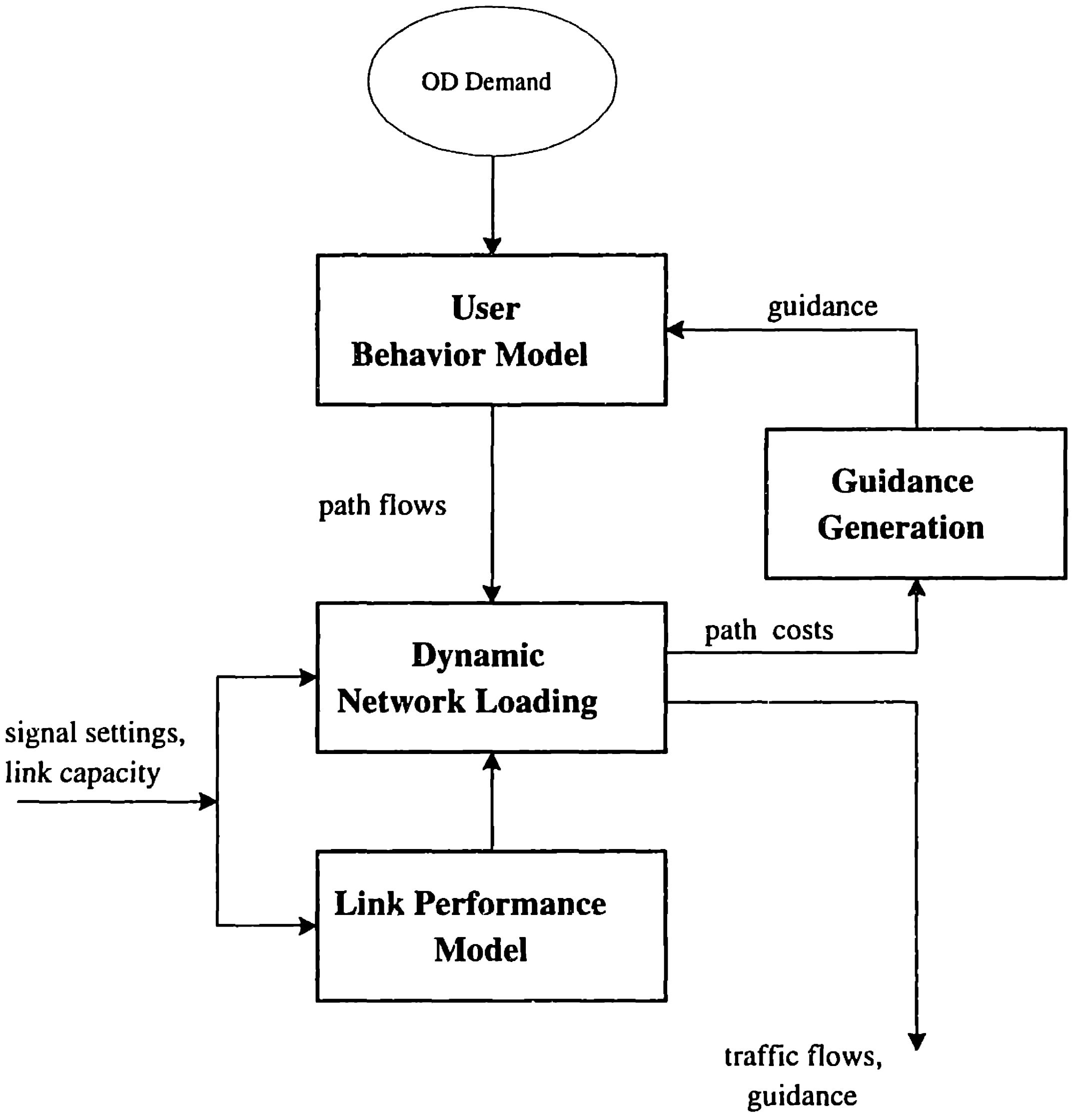

Figure 3-5: Dynamic Traffic Assignment (DTA) System

generation model produces predictive route guidance information for users based on current and future traffic conditions and in anticipation of users' responses. The DTA system enforces consistency between the traffic predictions on which the guidance is based and the traffic conditions that result from users' reactions to the guidance. Detailed descriptions of the above four components are provided next.

3.3.1 Guidance Generation

The guidance generation model's objective is to provide unbiased and consistent guidance information to users. Consistent and unbiased guidance implies that the traffic predictions (flows and travel costs) on which the guidance generation are based are unbiased estimates of conditions that users actually experience. The guidance generation's objective is in the users' best interest and therefore it eventually leads to user-optimality.

There are two types of travel guidance that a DTA system can provide: anticipatory descriptive guidance and prescriptive guidance. Anticipatory descriptive guidance gives users information on traffic conditions they are likely to encounter on various feasible paths from their current position to their destination. Anticipatory prescriptive guidance advises users of a path based on expected traffic conditions along alternative feasible paths.

3.3.2 Behavioral Model of Multiple User Classes

At the core of any traffic assignment model is modeling of user behavior. In the presence of an ITS travel guidance information, it becomes even more important to capture realistic behavior of heterogeneous users, particularly the users' information gathering, learning and complying process.

Four classes of users can be identified according to their information processing mechanism: habitual users, unguided users, imperfectly guided users and perfectly guided users.

Class 1: Habitual Users

The first class of users are those who habitually choose a fixed route. Class 1 users include those who have no any alternative other than to follow their scheduled paths (such as bus drivers), those who are unaware of other alternatives (such as unfamiliar tourists), or those who simply refuse to change their daily routing plans because they consider it an inconvenience to switch to different routes.

Regardless of the network condition, the route choices of class 1 users are fixed and therefore their *optimal* route choices are represented as follows:

$$h_{1p}^{rs}(t) = D_1^{rs}(t)\ \xi_p^{rs}, \qquad \forall r, s, p \tag{3.2}$$

where,

$h_{1p}^{rs}(t)$ = flow of class 1 users on path p departing from origin r to destination s at time t;

$D_1^{rs}(t)$ = demand flow rate of class 1 users for O-D pair (r,s) at time t:

$$\xi_p^{rs} = \begin{cases} 1 & \text{if } p \text{ is a habitual path from } r \text{ to } s \\ 0 & \text{otherwise} \end{cases}$$

Class 2: Unguided Users

The second class of users are those who do not have access to guidance information. The only source of information is past experience. They have neither accurate nor complete information about the network and their travel decisions are made to minimize *perceived travel time*, which is a random variable. Therefore, this class of users is characterized by their stochastic dynamic user-optimal (SDUO) behavior without guidance information.

The optimal route choice conditions for class 2 users are expressed as the following SDUO conditions:

$$P_{2p}^{rs}(t) = Prob\left[c_p^{rs}(t) + \epsilon_{2p}^{rs}(t) \ \le\ c_k^{rs}(t) + \epsilon_{2k}^{rs}(t) \ \mid \forall k \in \mathcal{P}^{rs}\right] \tag{3.3}$$

$$h_{2p}^{rs}(t) = D_2^{rs}(t)\, P_{2p}^{rs}(t), \quad \forall r, s, p \tag{3.4}$$

where,

$P^{rs}_{2p}(t)$ = class 2 users' probability of choosing path p departing from origin r to destination s at time t;

$h^{rs}_{2p}(t)$ = flow of class 2 users on path p departing from origin r to destination s at time t;

$D^{rs}_{2}(t)$ = demand flow rate of class 2 users for O-D pair (r, s) at time t;

$c^{rs}_{p}(t) + \epsilon^{rs}_{2p}(t)$ = perceived cost on path p departing from origin r to destination s at time t;

$c^{rs}_{p}(t)$ = actual travel cost on path p departing from origin r to destination s at time t;

$\epsilon^{rs}_{2p}(t)$ = the random error of the perceived cost on path p for class 2 users departing from origin r to destination s at time t.

Class 3: Imperfectly Guided Users

The third class of users are those who have access to real-time guidance information but do not comply with the guidance 100% or do not believe that the guidance information is 100% accurate. Their route choice behavior is similar to that of class 2 users. However, class 3 users' perceptions are more accurate than those of class 2 users since the former have extra information about traffic conditions. Therefore, class 3 users also follow stochastic dynamic user-optimal (SDUO) routes but the stochasticity is less than that of class 2 users.

The optimal route choice conditions for class 3 users are expressed as the following SDUO conditions:

$$P^{rs}_{3p}(t) = Prob\left[c^{rs}_{p}(t) + \epsilon^{rs}_{3p}(t) \;\le\; c^{rs}_{k}(t) + \epsilon^{rs}_{3k}(t) \;\mid \forall k \in \mathcal{P}^{rs}\right] \qquad (3.5)$$

$$h^{rs}_{3p}(t) = D^{rs}_{3}(t)\, P^{rs}_{3p}(t), \quad \forall r, s, p \qquad (3.6)$$

where,

$P_{3p}^{rs}(t)$ = class 3 users' probability of choosing path p departing from origin r to destination s at time t;

$h_{3p}^{rs}(t)$ = flow of class 3 users on path p departing from origin r to destination s at time t;

$D_3^{rs}(t)$ = demand flow rate of class 3 users for O-D pair $(r,\ s)$ at time t;

$\epsilon_{3p}^{rs}(t)$ = the random error of the perceived cost on path p for class 3 users departing from origin r to destination s at time t.

Since the perceived travel cost of class 3 users, $c_p^{rs}(t) + \epsilon_{3p}^{rs}(t)$, is more accurate than that of the class 2 users, $c_p^{rs}(t) + \epsilon_{3p}^{rs}(t)$, the variance of $\epsilon_{3p}^{rs}(t)$ is less than that of $\epsilon_{2p}^{rs}(t)$. That is,

$$Var\left[\epsilon_{3p}^{rs}(t)\right] \ \leq \ Var\left[\epsilon_{2p}^{rs}(t)\right] \tag{3.7}$$

Class 4: Perfectly Guided Users

The fourth and final class of users are those who have access to real-time guidance information and believe that the guidance information is 100% accurate. They fully comply with the guidance and their route choices are characterized as the deterministic dynamic user-optimum (DUO) since they only choose the dynamic shortest paths from their origins to their destinations.

The optimal route choice conditions for class 4 users are presented in the following DUO conditions:

$$\sum_p h_{4p}^{rs}(t) = D_4^{rs}(t), \quad \forall r, s, p \tag{3.8}$$

$$h_{4p}^{rs}(t)\left\{c_p^{rs}(t) - \pi^{rs}(t)\right\} = 0, \quad \forall r, s, p \tag{3.9}$$

$$c_p^{rs}(t) - \pi^{rs}(t) \geq 0, \quad \forall r, s, p \tag{3.10}$$

$$h_{4p}^{rs}(t) \geq 0, \quad \forall r, s, p \tag{3.11}$$

where,

$h_{3p}^{rs}(t)$ = flow of class 2 users on path p departing from origin r to destination s at time t;

$D_4^{rs}(t)$ = demand flow rate of class 4 users for O-D pair (r, s) at time t;

$\pi^{rs}(t)$ = $\min_p c_p^{rs}(t)$, minimum actual travel cost on path p departing from origin r to destination s at time t.

It should be noted that class 3 and class 4 users are special cases of class 2 users. The perception error of class 3 is smaller than that of class 2, while class 4 has a zero perception error.

The classification of users is not limited to the above four types. We can classify up to an infinite number of user classes based on the distribution of the random perception error. In general, the users' route choice behavior can be modeled using the following stochastic dynamic user-optimal (SDUO) conditions for multiple user classes:

$$P_{np}^{rs}(t) = Prob(c_p^{rs}(t) + \epsilon_{np}^{rs}(t) \leq c_k^{rs}(t) + \epsilon_k^{rs}(t), \forall k \in \mathcal{P}^{rs}) \tag{3.12}$$

$$h_{np}^{rs}(t) = D_n^{rs}(t)\ P_{np}^{rs}(t), \quad \forall r, s, p \tag{3.13}$$

where subscript n indicates the users of class n.

C-Logit Model for Stochastic Route Choice with Path Overlapping

Two discrete choice models—*Logit* and *Probit*—are available for calculating stochastic route choice probabilities based on random utility theory (see Ben-Akiva and Lerman, 1985 for a review on this subject). Both models have their pros and cons.

The Logit model assumes that the random error term, $\epsilon_{np}^{rs}(t)$, is i.i.d.[3] with a Gumble distribution. The Multinomial Logit (MNL) model for the stochastic route

[3] i.i.d. — independently and identically distributed.

choice is given by:

$$P_{np}^{rs}(t) = \frac{exp[-\alpha_n c_p^{rs}(t)]}{\sum_k exp[-\alpha_n c_k^{rs}(t)]} \tag{3.14}$$

where $P_{np}^{rs}(t)$ is the probability of class n users choosing path p of O-D pair (r, s) at time t and α_n is a parameter depending on the variance of $\epsilon_{np}^{rs}(t)$. The assumption of the random error's independent and identical distribution results in the property of independence of irrelevant alternative (IIA), which is not theoretically acceptable to the route choice problem when alternative paths share many common links. It has been suggested in the literature (Daganzo and Sheffi, 1997, and Sheffi, 1985) that the Logit model should be used in a network when strongly overlapping paths are excluded from the choice set. However, in an urban street network, the i.i.d. assumption of the Logit model can be easily violated.

The other well-known discrete choice model—*Probit*—is obtained if the random error terms are assumed to be jointly Normal distributed with mean zero. Unlike the MNL model, the Multinomial Probit (MNP) model does not have the undesirable IIA property since it does not require the random errors to be i.i.d. However, the application of a Probit model is limited by its computational difficulty because it is impossible to obtain an explicit functional form of the choice probability.

Recently, Cascetta *et al.* (1996) proposed a path choice model named C-Logit that overcomes the shortcoming of MNL–unrealistic choice probabilities in the presence of path overlapping. C-Logit is based on a different specification for the systematic utility of a Multinomial Logit model. C-Logit overcomes the undesirable IIA property of the Logit model while keeping the Logit model's closed analytical structure. The choice probability of the C-Logit model is expressed in the following form:

$$P_{np}^{rs}(t) = \frac{exp[-\alpha_n c_p^{rs}(t) - CF_p]}{\sum_k exp[-\alpha_n c_k^{rs}(t) - CF_k]} \tag{3.15}$$

where CF_p is a *commonality factor* for path p. It can be specified in different ways in order to account for the path overlapping. Cascetta *et al.* (1996) suggested several specifications for the commonality factor. In this thesis, the following commonality factor specification is used:

$$CF_p = \beta \; ln \left[\sum_k \left(\frac{L_{pk}}{\sqrt{L_p L_k}} \right)^{\gamma} \right] \tag{3.16}$$

where L_p and L_k are the lengths of paths p and k respectively and L_{pk} is the length of the common links shared by paths p and k. β and γ are two positive parameters. The link lengths can be measured by physical link lengths or better yet, the link-additive part of the generalized cost.

3.3.3 Dynamic Network Loading

The dynamic network loading (DNL) problem aims at finding time-dependent link flows, link travel costs and path travel costs, given route choices (path departure flows) and link performance models. The DNL problem essentially maps time-dependent path flows to the corresponding link variables. As reviewed in Chapter 2, the DNL problem has received considerable attention due to its importance in the dynamic traffic assignment (DTA) problem.

Coinciding with the two approaches to the DTA problem (see the review in Section 2.2), there have also been two approaches to modeling the dynamic network loading problem. One is simulation-based modeling, including CONTRAM (Leonard *et al.*, 1989), DYNASMART (Mahmassani *et al.*, 1992), MITTNS (Ben-Akiva *et al.*, 1994b), MITSIM (Yang and Koutsopoulos, 1996 and Yang, 1997) and DynaMIT (Ben-Akiva *et al.*, 1997). The advantage of the simulation approach is its detailed representation of network traffic movements at either a mesoscopic or even microscopic level.

Another approach to the DNL problem is based on flow-based analytical models (Friesz et al. 1993, Ran et al. 1993, Wu et al. 1995 and Xu et al. 1996). Recently, Chabini and He (1997, 1998a, 1998b) and He (1997) developed a general flow-based DNL formulation. Similar to those by Friesz et al. and Wu et al., the model is formulated as a system of equations representing the link dynamics, flow conservation, flow propagation and boundary constraints. Compared to previous models, Chabini and He's model has two major advantages:

1. the model is valid even if the FIFO condition is not satisfied; and

2. the model does not depend on any particular link performance function.

The mathematical representation of the DNL problem given below is drawn from Chabini and He (1998b):

Link dynamics equations:

$$\frac{dx^{rs}_{ap}(t)}{dt} = u^{rs}_{ap}(t) - v^{rs}_{ap}(t), \qquad \forall (r,s), \forall p \in \mathcal{P}^{rs}, \forall a, \tag{3.17}$$

Flow conservation equations:

$$u^{rs}_{ap}(t) = \begin{cases} h^{rs}_{p}(t), & a \in \mathcal{A}(r) \text{ (a is the first link on path p)} \\ v^{rs}_{a'p}(t), & a \notin \mathcal{A}(r) \text{ and a is after a'} \end{cases} \tag{3.18}$$

Flow propagation equations:

$$V^{rs}_{ap}(t) = \int_{w \in W} u^{rs}_{ap}(w)dw, \qquad \forall a, \forall (r,s), \forall p \in \mathcal{P}^{rs} \tag{3.19}$$

$$U^{rs}_{ap}(t) = U^{rs}_{ap}(0) + \int_0^t u^{rs}_{ap}(w)dw \qquad \forall a, \forall (r,s), \forall p \in \mathcal{P}^{rs} \tag{3.20}$$

$$V^{rs}_{ap}(t) = V^{rs}_{ap}(0) + \int_0^t v^{rs}_{ap}(w)dw \qquad \forall a, \forall (r,s), \forall p \in \mathcal{P}^{rs} \tag{3.21}$$

where $W = \{w : s_a(w) \leq t\}$ and $s_a(w)$ is the link exit time for a vehicle entering link a at time w.

Definitional constraints:

$$u_a(t) = \sum_{r,s,p} u^{rs}_{ap}(t), \qquad \forall a \tag{3.22}$$

$$v_a(t) = \sum_{r,s,p} v^{rs}_{ap}(t), \qquad \forall a \tag{3.23}$$

$$x_a(t) = \sum_{r,s,p} x^{rs}_{ap}(t), \qquad \forall a \tag{3.24}$$

Boundary conditions:

$$U_{ap}^{rs}(0) = 0, \quad V_{ap}^{rs}(0) = 0, \quad x_{ap}^{rs}(0) = 0, \qquad \forall (r,s), \forall p \in \mathcal{P}^{rs} \tag{3.25}$$

Non-negativity conditions:

$$x_{ap}^{rs}(t), u_{ap}^{rs}(t), v_{ap}^{rs}(t) \geq 0, \qquad \forall a, (r,s), \forall p \in \mathcal{P}^{rs} \tag{3.26}$$

See Appendix A for notations. In the above system of equations, the known variables are path departure flow rates $h_p^{rs}(t)$ and link travel time functions, $\tau_a(t)$, which are the outputs from the *link performance model.* The unknown variables are link inflow $u_{ap}^{rs}(t)$, link outflow $v_{ap}^{rs}(t)$, number of vehicles $x_{ap}^{rs}(t)$, cumulative link inflow $U_{ap}^{rs}(t)$, and cumulative link outflow $V_{ap}^{rs}(t)$.

The dynamic network loading component in our integrated framework (Figure 3-2) can be implemented as either a simulation-based model or a flow-based analytical model.

3.3.4 Link Performance Model

Link performance functions (or link delay functions) play a significant role in both a dynamic traffic assignment model and a dynamic traffic control model. In a DTA model, link performance functions are needed for the dynamic network loading model to generate link traffic conditions. In a DTC model, these functions are needed to estimate link delays and evaluate a control policy's performance.

The link performance function for interrupted traffic depends on both flow and control variables. Furthermore, it should be a function of the past system variables. That is,

$$\tau_a(t) = \tau_a[u(s), x(s), g(s) \mid s \leq t] \tag{3.27}$$

where $u(s), x(s), g(s)$ represent link flow, density (number of vehicles), and green split vectors at time s respectively. The functional form of $\tau_a(t)$ can be user-defined but the following requirements must be satisfied:

- τ_a is continuous and differentiable with respect to flow, density, and green split;
- τ_a is decreasing with respect to green split and increasing with respect to flow and density: $\frac{\partial \tau_a}{\partial g} \leq 0$, $\frac{\partial \tau_a}{\partial u} \geq 0$, $\frac{\partial \tau_a}{\partial x} \geq 0$.

The most popular link performance function in traffic signal control, however, is the Webster delay function (Webster, 1958). If vehicle arrivals at an intersection in *uniform*, the average delay can be calculated using queuing theory as follows:

$$d_a = \frac{\rho_a^2}{2u_a(1-\rho_a)} \tag{3.28}$$

where $\rho_a = \frac{u_a}{s_a g_a}$ is the saturation degree. Formula (3.28) often underestimates the average delay because it does not account for stochastic variation in the arrival rate. Webster (1958) added another term to the above formula in order to account for stochastic arrivals:[4]

$$\tau_a = \frac{9}{20}\left\{\frac{C_i[1-g_a]^2}{1-\frac{u_a}{s_a}} + \frac{(\rho_a)^2}{u_a[1-\rho_a]}\right\} \tag{3.29}$$

where C_i is the cycle time for intersection i (to which link a connects).

The *Highway Capacity Manual* (HCM, 1994) gives another delay function for estimating average delays at signalized intersections. Like the Webster delay function, the HCM model has two delay terms as well—one based on uniform arrivals; the other based on random arrivals. This HCM delay function is given by:

$$\tau_a = 0.38\, C_i \frac{[1-g_a]^2}{1-\frac{u_a}{s_a}} + 173\,(\rho_a)^2 \left[(\rho_a - 1) + \sqrt{(\rho_a-1)^2 + 16\frac{\rho_a}{s_a}}\right] \tag{3.30}$$

[4]The original Webster delay function is:

$$\tau_a = \frac{C_i[1-g_a]^2}{2[1-\frac{u_a}{s_a}]} + \frac{(\rho_a)^2}{2u_a[1-\rho_a]} - 0.65\left(\frac{C}{u^2}\right)^{1/3} \rho_a^{2+5g_a}$$

Since the last term is too complicated, Formula (3.29) is often used as an approximation.

3.3.5 Computation of Path Travel Costs

The link-based network conditions are determined by the dynamic network loading problem and the link performance model. However, path travel times are needed by the behavioral model to calculate users' route choice probabilities. The computation of path travel times is described as follows.

Consider path p consisting of sorted nodes $N_p^{rs} = (r, 1, 2, \cdots, i, \cdots, s-1, s)$ from r to s. When a user departs from origin r at time t, the travel time can be computed recursively as follows:

$$
\begin{aligned}
c_p^{r,1}(t) &= \tau_{r,1}(t) && (3.31)\\
c_p^{r,2}(t) &= c_p^{r,1}(t) + \tau_{1,2}(t + c_p^{r,1}(t)) && (3.32)\\
&\vdots \\
c_p^{r,i}(t) &= c_p^{r,i-1}(t) + \tau_{i-1,i}(t + c_p^{r,i-1}(t)) && (3.33)\\
&\vdots \\
c_p^{r,s}(t) &= c_p^{r,s-1}(t) + \tau_{s-1,s}(t + c_p^{r,s-1}(t)) && (3.34)
\end{aligned}
$$

where $\tau_{i,j}$ is the actual link travel time on link (i, j).

It is worth noting that the path travel time is the travel time experienced by the users. Path travel time depends not only on the network conditions at the time of departure, but also on future network conditions encountered while users are traveling.

Chapter 4

Formulations of the Combined Dynamic Traffic Control–Assignment Problem

This chapter proposes a game-theory methodology to study the combined dynamic traffic control-assignment (DTCA) problem and formulates a set of mathematical models based on different types of games between a traffic authority and highway users.

4.1 A Game Theory Methodology

The interaction between traffic control and assignment is essentially the conflict between two types of players: a traffic authority and highway users. The highway users minimize their own individual travel disutilities by choosing their routes while the traffic authority minimizes an overall system objective, such as the total delay, by setting signal timings. This conflict can be studied as a game between the users and the traffic authority.

Let $\Omega = \{\mathcal{N}, \mathcal{A}\}$ denote a traffic network, consisting of a set of intersection nodes $\mathcal{N}$ and a set of links $\mathcal{A}$. The problem for the users is to find dynamic user-optimal path flows over the network Ω, given a time-dependent O-D demand and signal timings.

The problem for the traffic authority is to find dynamic system-optimal signal timings, given time-dependent link flows.

Game theory has long been used to analyze individuals' economic behavior. See Kreps (1990), Fudenberg (1991) and Gibbons (1992) for a comprehensive review of game theory and its applications in economics. Since users' travel behavior is a subset of the general economic behavior, game theory can provide a powerful analytical tool to study the traffic assignment problem and its interaction with the traffic control. The traffic assignment problem is essentially a Nash non-cooperative game (Nash 1950; Nash 1951), while its interaction with traffic control fits within a Cournot game (Cournot 1838) or a Stackelberg game (von Stackelberg 1934).

This section reviews some fundamentals of game theory as applied to transportation problems and proposes a method to study the combined dynamic traffic control–assignment (DTCA) problem.

4.1.1 Nash Equilibria for Traffic Assignment

A traffic equilibrium often refers to the user equilibrium in which not a single user has incentive to change his or her travel decision. This user equilibrium is essentially a *Nash* equilibrium (Nash 1950; Nash 1951) for an n-person non-cooperative game among the users. To be more rigorous, let us define the Nash equilibrium. Suppose there are n players in the game and they are numbered from 1 to n. Let $\mathcal{H}_i$ denote the set of strategies available to player i (called i's strategy space), and h_i an arbitrary member of this set. Let $(h_1, \ldots, h_n)$ denote a combination of strategies, one for each player, and $Z_i(h_1, \ldots, h_n)$ player i's payoff function when the players choose the strategies $(h_1, \ldots, h_n)$. The *normal form* of the game is denoted as $\Lambda = \{\mathcal{H}_1, \ldots, \mathcal{H}_n; Z_1, \ldots, Z_n\}$. In the case of a traffic assignment, each player is a user; the strategy space $\mathcal{H}_i$ is the set of all available paths from player i's origin to his/her destination and the payoff function Z_i is the travel utility for player i.

Figure 4-1 gives the *extensive form* of an n-player game for a route choice problem. Each user i chooses one path out of his/her feasible set $\mathcal{H}_i$ independently. The users choose their paths as if they happen in a sequence, but when user i makes a decision,

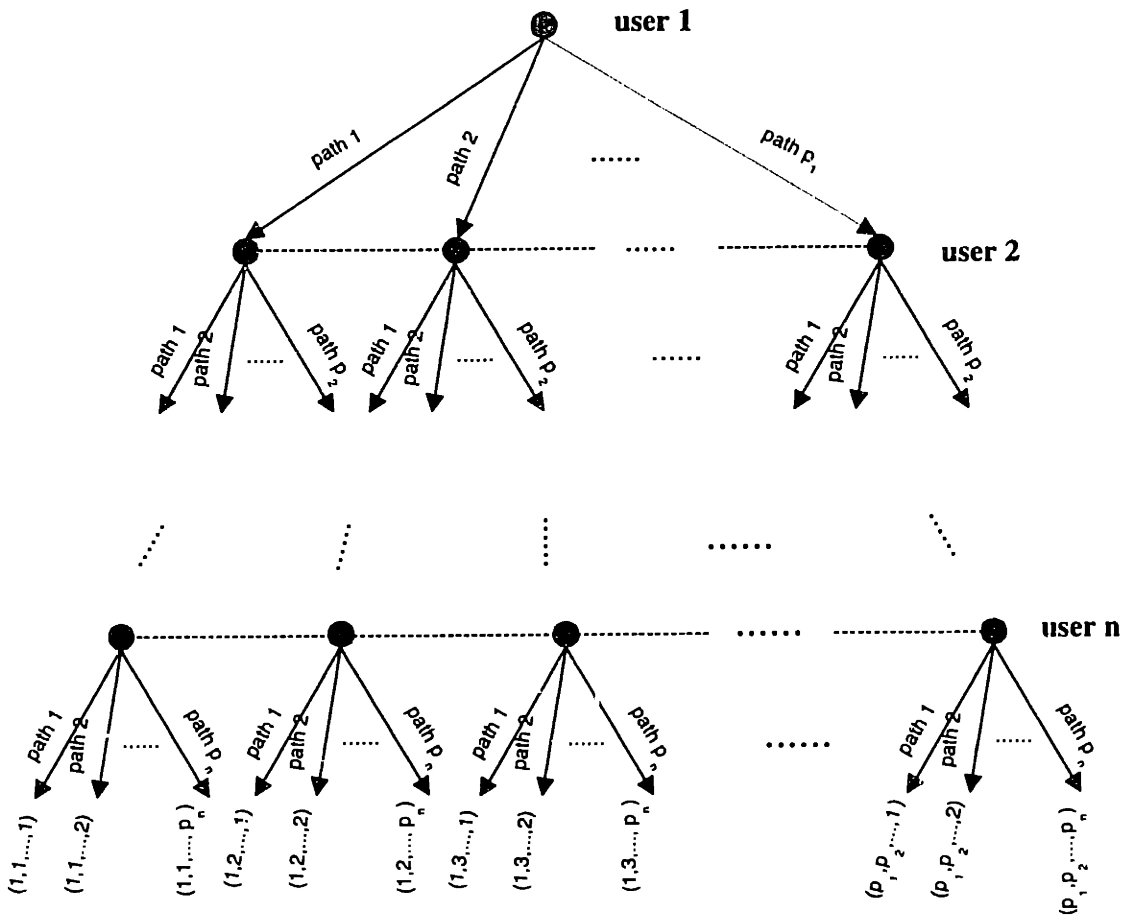

Figure 4-1: A Non-Cooperative n-Player Game Among Users

he/she does not know the others' selected paths. The dashed line that connects the nodes at each user level indicates that those nodes are indistinguishable. Each leaf at the end of the tree represents a possible outcome of the game.

The Nash equilibrium of the game is defined as follows:

Definition 1 (Nash Equilibrium) *In an n-player normal form game $\Lambda = \{\mathcal{H}_1, \dots, \mathcal{H}_n; Z_1, \dots, Z_n\}$, the combination of strategies $(h_1^*, \dots, h_n^*)$ is a Nash equilibrium if, for each player i, h_i^* is player i's best response to the strategies given by the other n-1 players, $(h_1^*, \dots, h_{i-1}^*, h_{i+1}^* \dots, h_n^*)$:*

$$Z_i(h_1^*, \dots, h_i^*, \dots, h_n^*) \geq Z_i(h_1^*, \dots, h_i, \dots, h_n^*), \qquad \forall h_i \in \mathcal{H}_i$$

That is, each player's strategy h_i^* solves

$$\max_{h_i \in \mathcal{H}_i} Z_i(h_1^*, \dots, h_i, \dots, h_n^*)$$

In the traffic assignment game, an assignment is a Nash equilibrium if no single user can improve his/her payoff function by unilaterally changing his/her route. Different assumptions about users' payoff functions determine the outcome of the Nash equilibrium. For example, if the users are assumed to have perfect information and their payoff functions are the negative travel costs, then the Nash equilibrium is a *pure strategy* and it is equivalent to a Wardropian equilibrium. If, however, the users do not know the actual travel costs and the payoff functions are the negative perceived travel costs, then the Nash equilibrium is a *mixed strategy* and it is equivalent to a stochastic user equilibrium (SUE).

4.1.2 Cournot Equilibria for One-Shot Games

In the combined control-assignment problem, there is another player—the traffic authority. The strategy space for the traffic authority is the feasible set of available control parameters and the payoff function is related to some overall system performance, such as the total delay. Adding the traffic authority to the game is not as simple as extending an n-player game to an $n+1$-player game, because the strategy space and the payoff function for this additional player differ from the rest of the n players. Rather, the combined control-assignment problem should be interpreted as the following two games being played: The first one is an n-player non-cooperative game among the users and the second one is a game between the traffic authority and the collective users. The first game is identical to the traffic assignment game defined above in Section 4.1.1. We assume that the outcome of the first game is a Nash equilibrium.

Our focus in this thesis is to investigate the second game—the game between the traffic authority and the collective users. The collective users are treated as one player who distributes the flows to the network so that the flow pattern is a Nash equilibrium among the users. The traffic authority chooses the control variables to achieve a system-level payoff.

The strategy space for the traffic authority includes all the feasible signal timings, g, for each signalized intersection. The payoff function for the traffic authority depends on the system objective, which can be a minimization of the total delay, a maximization of the capacity or other objective. In general, the traffic authority's payoff is a function of both traffic flows and green time splits. Denote its payoff function as $Z_g(g, h)$. The strategy space for the users, $\mathcal{H}$, is all the feasible traffic flows for the given network and demand. The payoff function for the users is $Z_h(g, h)$, which depends on both flow and control. Under the above settings, the game for the combined control-assignment problem can be denoted as $\Lambda = \{\mathcal{G}, \mathcal{H}; Z_g(g, h), Z_h(g, h)\}$.

Figure 4-2 shows the game between the traffic authority and the collective users. This type of a duopoly game, in which two players choose their strategies simultaneously, is known as a *Cournot game* (Cournot 1838). It is a one-shot game. The equilibrium of this Cournot game is defined as follows:

Definition 2 (Cournot Equilibrium) *In the two-player game between the traffic authority and the users, $\Lambda = \{\mathcal{G}, \mathcal{H}; Z_g(g, h), Z_h(g, h)\}$, the combination of strategies (g^*, h^*) is a Cournot equilibrium if control plan g^* is the traffic authority's best*

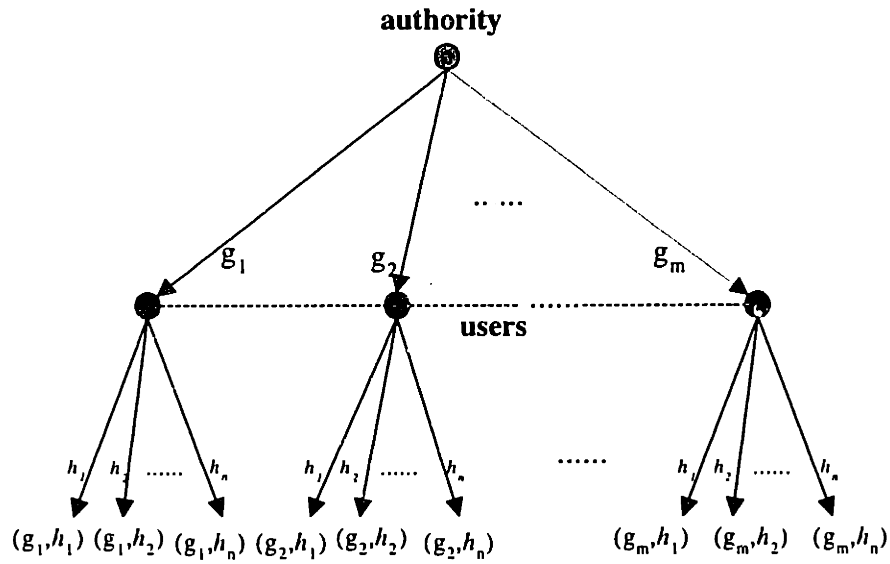

Figure 4-2: A Cournot game between a traffic authority and users

response to flow h^* and h^* is the users' best response to control plan g^*:

$$Z_g(g^*, h^*) \geq Z_g(g, h^*) \ \ and \ \ Z_h(g^*, h^*) \geq Z_h(g^*, h), \qquad \forall g \in \mathcal{G}, \ h \in \mathcal{H}$$

That is, strategy combination (g^*, h^*) solves the following two problems simultaneously:

$$\max_{g \in \mathcal{G}} Z_g(g, h^*) \ \ and \ \ \max_{h \in \mathcal{H}} Z_h(g^*, h)$$

4.1.3 Stackelberg Equilibria for Bi-Level Games

In a one-shot game such as a Cournot game, all participating players move simultaneously and therefore one player's response is unknown in advance to others. However, if we assume that the users' route choice behavior characterized by Nash equilibrium is common knowledge, the traffic authority can design a better strategy taking the anticipated users' reactions into account.

In this section, we treat the combined control-assignment problem as the following two-stage game. First, the traffic authority sets the signal timings. Second, the users observe the traffic authority's signal settings and choose their best routes accordingly. Figure 4-3 gives the extensive form of this game. In the case of duopoly (two players), such a multi-stage game is known in economics as a Stackelberg game. The key characteristics of a multi-stage game of complete information are:

1. the moves occur in sequence;
2. all previous moves are observed before the next move is chosen;
3. the players' payoff from each feasible combination of moves are common knowledge.

The player who moves first is a leader and the player who moves next is a follower. In general, the equilibrium of the game is *not* a Nash equilibrium. Rather, the equilibrium is determined by *backwards induction*, as shown in Figure 4-3. The traffic authority first initiates the move by setting the control, g. When the users make their move at the game's second stage they will face the following problem:

$$\max_{h \in \mathcal{H}} \; Z_h(g, h) \tag{4.1}$$

Assume that for each g in $\mathcal{G}$, above optimization problem has a unique solution. Denote the unique optimal solution by $h^*(g)$. $h^*(g)$ is the users' reaction or best response to the traffic authority's strategy g. Since the traffic authority can also solve the users' optimization problem of (4.1), the traffic authority should anticipate the users' reaction to each possible control strategy g. Therefore the traffic authority's problem amounts to

$$\max_{g \in \mathcal{G}} \; Z_g(g, h^*(g)) \tag{4.2}$$

This type of a two-stage game is known as a Stackelberg game (von Stackelberg 1934), or a leader-follower game. The equilibrium of the Stackelberg game can thus be defined as:

Definition 3 (Stackelberg Equilibrium) *In the two-player game between the traffic authority and the users, $\Lambda = \{\mathcal{G}, H; Z_g, Z_h\}$, the combination of strategies (g^*, h^*) is a Stackelberg equilibrium if strategy combination (g^*, h^*) solves the following bi-level problem:*

$$\begin{aligned} &\max_{g \in \mathcal{G}} \;\; Z_g(g, h^*(g)) \\ &s.t. \;\; h^*(g) = \max_{h \in \mathcal{H}} \; Z_h(g, h) \end{aligned} \tag{4.3}$$

Compared to the payoff in a Cournot equilibrium, the payoff for the traffic authority in a Stackelberg equilibrium should be higher (or at least equal) since the users' response has been taken into account when the control strategy is selected. This will be verified later in the detailed formulations of the one-shot and the bi-levei games.

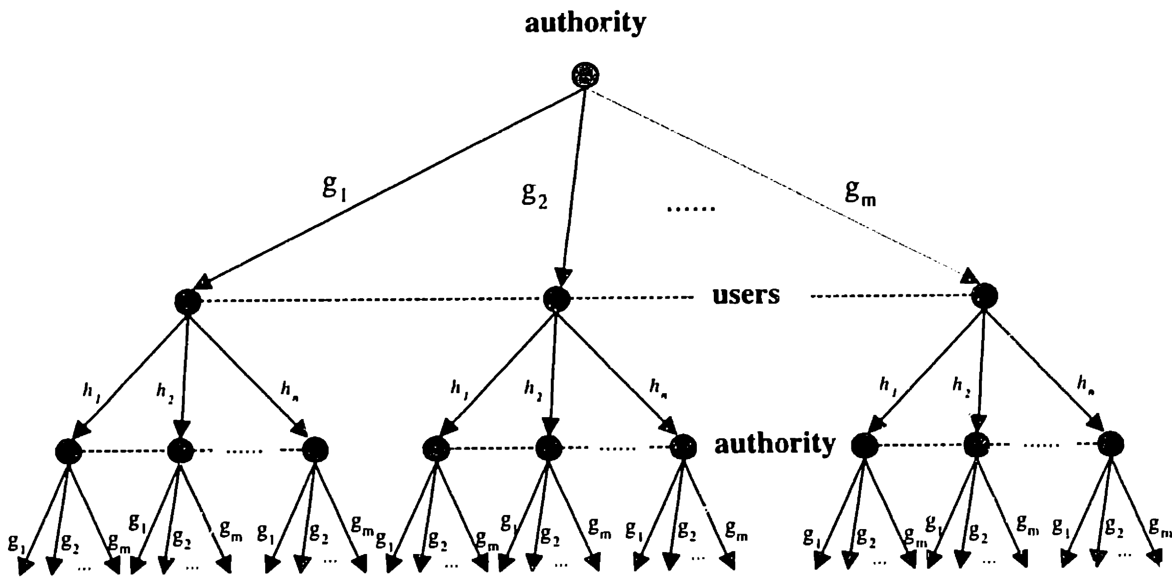

Figure 4-3: A Stackelberg game between a traffic authority and users

4.2 Dynamic Traffic Control Problem

This section formulates the dynamic traffic control (DTC) problem and discusses some of its important properties.

4.2.1 Formulation of the Dynamic Traffic Control Problem

Assume that the phasing and cycle time for each intersection are given. The problem for the traffic authority is to allocate a green time for each phase. The problem is formulated as follows:

(P1)

$$\min_{G} Z_G = \int_0^T \sum_{a \in \mathcal{A}} u_a(t)\ \tau_a[u_a(t), x_a(t), G_a(t)]\, dt \tag{4.4}$$

st.

$$\sum_{m} [G_i^m(t) + l_i^m] = C_i(t), \qquad \forall i, m \tag{4.5}$$

$$G_a(t) = \sum_{m \in \mathcal{M}_i} G_i^m(t)\ \theta_{ia}^m(t), \qquad \forall i, a \in B(i) \tag{4.6}$$

$$G_i^m(t) \geq G_i^{m,min} \qquad \forall i, m \tag{4.7}$$

where[1]

$G_a(t)$	=	green time for link a at time t (variable);
$G_i^m(t)$	=	green time phase m of intersection i at time t (variable);
$\tau_a(t)$	=	$\tau_a[u_a(t), x_a(t), G_a(t)]$ link delay function (funtional form given);
$u_a(t)$	=	entry flow on link a at time t (given);
$x_a(t)$	=	number of vehicles on link a at time t (given);
$l_i^m(t)$	=	lost time for phase m of intersection i at time t (given);
$C_i(t)$	=	cycle time for intersection i at time t (given);
$G_i^{m,min}$	=	minimum green time for link a at time t (given);
T	=	the total length of the time horizon (given);

[1] See Appendix A for all notations used in this thesis

$$\theta_{ia}^m(t) \quad = \begin{cases} 1 & \text{link } a \text{ belongs to phase } m \text{ at time } t \\ 0 & \text{otherwise} \end{cases} \quad \text{(given)}.$$

The objective is to minimize the total delay in the network. Constraint (4.5) states that the summation of the green time and the lost time over all phases of an intersection equals the cycle time of that intersection. Constraint (4.6) defines the relationship between a link green time, $G_a(t)$, and a phase green time, $G_i^m(t)$. Equation (4.7) is a minimum green time constraint for each phase.

$\tau_a(t)$ is a link performance function for interrupted traffic that depends on both flow and control variables. Section 3.3.4 discusses some regularity requirements and some common functional forms of the link performance function.

The above mathematical programming (MP) problem (P1) can be simplified by normalizing the decision variable of a green time, $G_i^m(t)$, to a green split, $g_i^m(t)$, as follows:

$$\begin{aligned} g_i^m(t) &\equiv \frac{G_i^m(t) - G_i^{m,min}}{C_i(t) - \sum_m (l_i^m + G_i^{m,min})} \\ &= \frac{G_i^m(t) - G_i^{m,min}}{\sum_m (G_i^m(t) - G_i^{m,min})} \end{aligned} \tag{4.8}$$

Here split $g_i^m(t)$ can be interpreted as the proportion of the allocated green time beyond the minimum, to the overall allocatable capacity for the intersection. Program (P1) is then standardized to:

(P2)

$$\min_g Z_g = \int_0^T \sum_{a \in A} u_a(t)\, \tau_a[u_a(t), x_a(t), g_a(t)]dt \tag{4.9}$$

$st.$

$$\sum_m g_i^m(t) = 1, \qquad \forall i, m \tag{4.10}$$

$$g_a(t) = \sum_{m \in \mathcal{M}_i} g_i^m(t)\, \theta_{ia}^m(t), \qquad \forall i, a \in B(i) \tag{4.11}$$

$$g_i^m(t) \geq 0, \qquad \forall i, m \tag{4.12}$$

where constraints (4.10)–(4.12) correspond exactly to constraints (4.5)–(4.7). We denote the feasible set defined by constraints (4.10)–(4.12) as $\mathcal{G}$.

4.2.2 Dynamic System Optimum

Program (P2) is a special case of an optimal control problem. Program (P2) contains no state variable—it only has control variables (g's) and input variables (u's and x's). In order to derive the optimality conditions, we define the following Hamiltonian function:

$$H = \sum_{a \in \mathcal{A}} u_a(t)\tau_a(t) + \sum_i \pi_i(t)(1 - \sum_m g_i^m(t)) \tag{4.13}$$

where $g_i^m(t) \geq 0$ and $\pi_i(t)$ is a Lagrange multiplier associated with constraint (4.10). The first order necessary conditions for Program (P2) can be derived as follows:

$$g_i^m(t)\,\frac{\partial H}{\partial g_i^m(t)} = 0; \qquad \forall i, m \tag{4.14}$$

$$\frac{\partial H}{\partial g_i^m(t)} \geq 0; \qquad \forall i, m \tag{4.15}$$

$$g_i^m(t) \geq 0; \qquad \forall i, m \tag{4.16}$$

If we define the *marginal link delay*, $\tilde{\tau}_a(t)$, as the partial derivative of the total link delay with respect to link green time split:

$$\tilde{\tau}_a(t) = \frac{\partial[u_a(t)\tau_a(t)]}{\partial g_a(t)} = u_a(t)\,\frac{\partial \tau_a(t)}{\partial g_a(t)} \tag{4.17}$$

then the marginal phase delay, $\tilde{c}_i^m(t)$, can be obtained as:

$$\tilde{c}_i^m(t) = \sum_a \tilde{\tau}_a(t)\,\theta_{ia}^m(t) \tag{4.18}$$

Since the link delay, $\tau_a(t)$, is a decreasing function of the link green time split, $g_a(t)$, both the marginal link delay and the marginal phase delay must be non-positive, i.e., $\tilde{\tau}_a(t) \leq 0$ and $\tilde{c}_i^m(t) \leq 0$. The partial derivative of the Lagrangian with respect to the green split, $\frac{\partial H}{\partial g_i^m(t)}$ in first order conditions of (4.14)–(4.16), is exactly the marginal phase delay $\tilde{c}_i^m(t)$. To see this, the derivative of the Lagrangian is derived as follows:

$$\begin{aligned}
\frac{\partial H}{\partial g_i^m(t)} &= \frac{\partial}{\partial g_i^m(t)} \sum_a u_a(t)\tau_a(t) - \pi_i(t) \\
&= \sum_a u_a(t) \frac{\partial \tau_a(t)}{\partial g_a(t)} \frac{\partial g_a(t)}{\partial g_i^m(t)} \\
&= \sum_a u_a(t) \frac{\partial \tau_a(t)}{\partial g_i^m(t)} \cdot \theta_{ia}^m \\
&= \sum_a \tilde{\tau}_a(t)\, \theta_{ia}^m \\
&= \tilde{c}_i^m(t)
\end{aligned} \tag{4.19}$$

Therefore, the first order conditions of (4.14)–(4.16) can be rewritten using the marginal phase delay as follows:

$$g_i^m(t)(\tilde{c}_i^m(t) - \pi_i(t)) = 0; \qquad \forall i, m \tag{4.20}$$

$$\tilde{c}_i^m(t) - \pi_i(t) \geq 0; \qquad \forall i, m \tag{4.21}$$

$$g_i^m(t) \geq 0; \qquad \forall i, m \tag{4.22}$$

From Equation (4.21), it follows immediately that

$$\pi_i(t) \leq \tilde{c}_i^m(t) \tag{4.23}$$

$$\pi_i(t) = \min_{m \in \mathcal{M}_i} \tilde{c}_i^m(t) \tag{4.24}$$

Therefore, the Lagrange multiplier $\pi_i(t)$ is the minimal marginal phase delay for intersection i at time t. The optimality conditions of (4.20)–(4.22) can thus be interpreted as follows:

Dynamic System-Optimum (DSO) Control: *For each intersection, any phase with a positive green time split must have an equal and the minimal marginal delay. Equivalently, for each intersection, green time split can only be assigned to a* *phase with the minimal marginal delay.*

4.2.3 Variational Inequality Problem for DTC

The equivalent variational inequality (VI) formulation of the DSO signal setting conditions (4.20)–(4.21) can be stated as follows (see Nagurney (1993) for a review on VI):

The VI Problem for DTC: The dynamic signal setting, $g_i^{m^*}(t)$, satisfying the timing constraint set (4.10)–(4.12) is a dynamic system optimal (DSO) setting if and only if it satisfies the variational inequality (VI) problem:

(P3)

$$\int_0^T \sum_i \sum_m \tilde{c}_i^{m^*}(t)(g_i^m(t) - g_i^{m^*}(t))dt \geq 0; \qquad \forall g_i^m(t) \in \mathcal{G} \tag{4.25}$$

where $\tilde{c}_i^{m^*}(t)$ is the marginal phase delay when the timing is $g_i^{m^*}(t)$. We denote the above VI problem as $\mathbf{VI}[\tilde{c}, \mathcal{G}]$. Notice that the marginal delay vector $\tilde{c}$ is a function of both the flow vector h and the green split vector g because the link cost τ is a function of both h and g [Equation (3.27)]. Therefore, we can also denote the VI problem of (P3) as $\mathbf{VI}[\tilde{c}(g,h), g \in \mathcal{G}]$.

It can be proved that any solution of the VI problem (P3) will satisfy the DSO conditions of (4.20)–(4.21) *and* any green time split satisfying the DSO conditions of (4.20)–(4.21) is a solution to the VI problem (P3). The VI formulation of (P3) is also equivalent to the mathematical programming (MP) formulation of (P2) since $\tilde{c}_i^{m^*}(t)$, by definition, is the gradient of the objective function Z_g, and Z_g is convex. However, the VI formulation of (P3) is in general broader than the MP formulation of (P2).

The existence of a solution to the VI problem, $\mathbf{VI}[\tilde{c}(g,h), g \in \mathcal{G}]$, is guaranteed as long as the feasible set $\mathcal{G}$ is closed and compact and the marginal delay function $\tilde{c}$ is continuous. The uniqueness of the VI problem of (P3) solution is also guaranteed if set $\mathcal{G}$ is closed and compact, and the marginal delay function $\tilde{c}$ is strictly monotone.

Clearly, set $\mathcal{G}$ defined by (4.10)–(4.12) is convex and therefore closed and compact. The property of a marginal delay function depends on the link delay function being used. For most link delay functions, their marginal delay functions are continuous and monotone.

4.3 Dynamic Traffic Assignment Problem

The game for the users is to find dynamic user-optimal path flows over the network, given time-dependent O-D demand flows and signal settings. This so-called DTA problem has been studied intensively in the literature. See the DTA literature review in Section 2.2. The DTA formulation given below is similar to the one developed by Chabini and He (1998a, 1998b).

The payoff function for each user is his or her travel utility function. The users choose their strategies of travel decisions (e.g., route choices) in order to maximize their payoff functions. Due to the heterogeneity of users, different users may have different functional forms of their payoff or utility functions. Based on the mechanism that users access and process information, multiple user classes are proposed in Section 3.3.2 to capture the hetergeneity in the users' decision process:

- Class 1: Habitual Users
- Class 2: Unguided Users
- Class 3: Imperfectly Guided Users
- Class 4: Perfectly Guided Users

It should be noted that the behavioral models of class 3 and class 4 are special cases of class 2's stochastic behavior model. Class 3 has a smaller perception error than that of class 2, while class 4 has a zero perception error. The classification of users is not limited to the above four types. We can classify up to an infinite number of user classes based on the distribution of the random perception error.

The stochastic optimal route choice conditions for class n users can be expressed as the following SDUO conditions:

$$P_{np}^{rs}(t) = Prob\left[c_p^{rs}(t) + \epsilon_{np}^{rs}(t) \;\leq\; c_k^{rs}(t) + \epsilon_{nk}^{rs}(t) \;\mid \forall k \in \mathcal{P}^{rs}\right] \tag{4.26}$$

$$h_{np}^{rs}(t) = D_n^{rs}(t)\, P_{np}^{rs}(t), \quad \forall r, s, p \tag{4.27}$$

where,

$P_{np}^{rs}(t)$ = the probability of choosing path p of O-D pair $(r,\ s)$ for class n users departing from origin r to destination s at time t;

$h_{np}^{rs}(t)$ = flow of class n users on path p departing from origin r to destination s at time t;

$D_n^{rs}(t)$ = demand flow rate of class n users for O-D pair $(r,\ s)$ at time t;

$c_p^{rs}(t) + \epsilon_{np}^{rs}(t)$ = perceived cost on path p of O-D pair $(r,\ s)$ for class n users departing from origin r at time t;

$c_p^{rs}(t)$ = actual travel cost on path p of O-D pair $(r,\ s)$ for class n users departing from origin r at time t;

$\epsilon_{np}^{rs}(t)$ = the random error of the perceived cost on path p of O-D pair $(r,\ s)$ for class n users departing from origin r at time t.

The optimal route choice conditions of equations (4.26)-(4.27) can be interpreted as follows:

Stochastic Dynamic User Optimum (SDUO) Assignment: *For each O-D pair of each user class, any path with a positive flow must have an equal and the minimal perceived path cost. In other words, an O-D flow of a user class can only be assigned to the dynamic perceived shortest path of that O-D pair.*

In order to construct an equivalent variational inequality (VI) formulation of the stochastic route choice problem, we define an auxiliary path cost function of each user class as follows:

$$E_{np}^{rs}(t) = \left[h_{np}^{rs}(t) - D_n^{rs}(t)\, P_{np}^{rs}(t)\right] \frac{\partial c_p^{rs}(t)}{\partial h_{np}^{rs}(t)} \tag{4.28}$$

where $\frac{\partial c_p^{rs}(t)}{\partial h_{np}^{rs}(t)}$ is the marginal path cost with respect to the path departure flow of class n users. It is assumed that the actual path cost,$c_p^{rs}(t)$, is a strictly increasing function of the path departure flow of a user class, $h_{np}^{rs}(t)$. That is,

$$\frac{\partial c_p^{rs}(t)}{\partial h_{np}^{rs}(t)} > 0 \tag{4.29}$$

The equivalent variational inequality (VI) problem of stochastic dynamic user optimal conditions (4.26)–(4.27) for class n users can be formulated as follows:

(P4)

$$\int_0^T \sum_{r,s} \sum_p E_{np}^{rs^*}(t)\ \left[h_{np}^{rs}(t) - h_{np}^{rs^*}(t)\right] dt\ \geq\ 0; \qquad \forall h_{np}^{rs}(t) \in \mathcal{H}_n \tag{4.30}$$

where $\mathcal{H}_n$ is the set of all feasible path flows for class n users and the asterisk denotes the path flows and auxiliary costs that correspond to the optimal solution.

Since Equation (4.30) holds true for *every* user class n, the equivalent VI problem for *all* user classes is then given by

(P5)

$$\int_0^T \sum_n \sum_{r,s} \sum_p E_{np}^{rs^*}(t)\ \left[h_{np}^{rs}(t) - h_{np}^{rs^*}(t)\right] dt\ \geq\ 0; \qquad \forall h_{np}^{rs}(t) \in \mathcal{H} \tag{4.31}$$

Feasible set $\mathcal{H}$ is the set of all feasible path flows for all users and it is defined by the dynamic network loading (DNL) constraints (Section 3.3.3), the link performance model (Section 3.3.4) and the equations for path cost computations (Section 3.3.5).

For convenience, we denote the VI problem (P5) as $\mathbf{VI}[E(g,h), h \in \mathcal{H}]$, in which $E(g,h)$ is defined in Equation (4.28), depending on both the green split g and the flow h.

4.4 Combining Control and Assignment: A Non-Cooperative Game

The interaction between control and assignment is essentially a non-cooperative game between two players: the traffic authority and the highway users. The traffic authority seeks a dynamic system-optimal (DSO) outcome by controlling signal timings and the users seek a stochastic dynamic user-optimal (SDUO) outcome by choosing their travel paths. Three types of the combined dynamic traffic control-assignment problems are formulated. First, the combined control-assignment problem is formulated as a one-level program—a Cournot game (Cournot 1838). Then the combined control-assignment problem is formulated as a bi-level program–a Stackelberg game (von Stackelberg 1934). Finally, a system optimal control-assignment problem is formulated as a monopoly game.

4.4.1 One-Level Control–Assignment Problem: A Cournot Game

The DTC and the DTA problems formulated in Sections 4.2 and 4.3 are first integrated as a one-level control-assignment problem, in which two sub-problems—DTA and DTC—are solved simultaneously. Such a simultaneous game between the traffic authority and the users is a Nash non-cooperative game (Nash 1950). In the case of two players, such a duopoly game is also known as a Cournot game (Cournot 1838). Figure 4-2 shows the *extensive form* of this Cournot game. It is a one-shot game in which each player makes his or her move independently, unaware of the other's strategy. Figure 4-2 shows an example where there are m possible strategies for the authority and n strategies for the users. Each combination of strategies determines the game's outcome. The equilibrium of the Cournot game is defined as follows:

Cournot Equilibrium: *In the non-cooperative game between the traffic authority and the highway users, the combination of strategies (g^*, h^*) is a Cournot equilibrium if and only if control plan g^* is the traffic authority's best response to flow h^* and* *flow h^* is the users' best response to control plan g^*.*

That is, the strategy combination (g^*, h^*) is a Cournot equilibrium if and only if it solves the following DTC and DTA problems simultaneously:

(P6)

DTC Problem:

$$\min_g Z_g = \int_0^T \sum_n \sum_{r,s} \sum_p c_p^{rs}(t)\, h_{np}^{rs^*}(t)\, dt \tag{4.32}$$

$$st. \qquad g_i^m(t) \in \mathcal{G}$$

DTA Problem:

$$\int_0^T \sum_n \sum_{r,s} \sum_p E_{np}^{rs^*}(t) \;\; \left[h_{np}^{rs}(t) - h_{np}^{rs^*}(t)\right] \, dt \;\geq\; 0; \qquad \forall h_{np}^{rs}(t) \in \mathcal{H} \tag{4.33}$$

Set $\mathcal{G}$ is defined by Equations (4.10)–(4.12 in Program (P2). Set $\mathcal{H}$ is the set of all feasible path flows for all users and is defined by the dynamic network loading (DNL) constraints (Section 3.3.3), the link performance model (Section 3.3.4) and the equations for path cost computaitons (Section 3.3.5).

The Cournot game formulated above can also be transformed to the following single variational inequality problem:

(P7)

$$\int_0^T \sum_{i,m} \tilde{c}_i^{m*}(t)[g_i^m(t) - g_i^{m*}(t)]\, dt + \int_0^T \sum_{p,r,s} E_{np}^{rs*}(t)\, [h_{np}^{rs}(t) - h_{np}^{rs^*}(t)]\, dt \geq 0$$

$$\forall g_i^m(t) \in \mathcal{G},\;\; h_{np}^{rs}(t) \in \mathcal{H} \tag{4.34}$$

where the first part of the left hand side is the DTC problem of $\mathbf{VI}[\tilde{c}(g,h), g \in \mathcal{G}]$ and the second part is the DTA problem of $\mathbf{VI}[E(g,h), h \in \mathcal{H}]$. The asterisk indicates the Cournot equilibrium solution.

The Cournot game defined in problem (P6) can also be represented by using a *fixed point* formulation. That is, solution (g^*, h^*) is a Cournot equilibrium if it solves the following two fixed point problems simultaneously:

(P8)

DTC Problem:

$$g = F_g(g, h^*) \qquad (h^* \text{ is given})$$

DTA Problem:

$$h = F_h(g^*, h) \qquad (g^* \text{ is given})$$

F_g is a mapping that gives the best control solution g, for a given traffic flow h. Likewise, F_h is a mapping that gives the best route choice solution for a given control. For a discussion of the relationships among mathematical programming, variational inequality and fixed point formulations, see Nagurney (1993).

The game between the traffic authority and the users is *not* a *zero-sum game.* This is true for both the Cournot game defined above and the Stackelberg game defined in the next section.

4.4.2 Bi-Level Control–Assignment Problem: A Stackelberg Game

Assume that the traffic authority knows in advance the users' optimal strategy characterized by the SDUO conditions of (4.26)–(4.27). The traffic authority can then take the anticipated users' reactions into account in designing the optimal control strategy.

We formulate the combined control-assignment problem as the following two-stage game. First, the traffic authority sets signal timings. Second, the users observe the signal timings and choose their best routes accordingly. Figure 4-3 gives the extensive form of this two-stage game. In the case of duopoly (two players), such a multi-stage dynamic game is known as a Stackelberg game (von Stackelberg 1934).

The player who moves first is a leader and the player who moves second is a follower. In general, the equilibrium of the game is *not* a Cournot equilibrium. Instead, the equilibrium is determined by backwards induction, as shown in Figure 4-3. The traffic authority first initiates the move by setting the control, g. When the users make their move at the second stage of the game they will choose the optimal flow h to solve the DTA problem of $\mathbf{VI}[E(g,h), h \in \mathcal{H}]$. For each given signal setting $g \in \mathcal{G}$, the optimal solution of the DTA problem is denoted by $h^*(g)$:

$$h^*(g) = \mathbf{VI}[E(g,h), h \in \mathcal{H}] \tag{4.35}$$

This optimal solution $h^*(g)$ is the users' reaction or best response to the traffic authority's strategy g. Since the traffic authority can solve the users' optimization problem of (4.35), the traffic authority should anticipate users' reaction to each possible control strategy g. Therefore the traffic authority's DTC problem amounts to

$$\min_{g \in \mathcal{G}} \; Z_g(g, h^*(g)) \tag{4.36}$$

where the objective function Z_g is defined in Equation (4.9) and the feasible set G is defined by Equations (4.10)–(4.12). The equilibrium of such a Stackelberg game can thus be defined as follows:

Stackelberg Equilibrium: *In the non-cooperative game between the traffic authority and the users, the strategy combination (g^*, h^*) is a Stackelberg equilibrium if and only if it solves the following bi-level programming problem:*

(P9)

$$\min_g Z_g = \int_0^T \sum_n \sum_{r,s} \sum_p c_p^{rs}(t)\, h_{np}^{rs^*}(t)\, dt \tag{4.37}$$

$$st. \quad g_i^m(t) \in \mathcal{G} \tag{4.38}$$

$$\int_0^T \sum_n \sum_{r,s} \sum_p E_{np}^{rs^*}(t) \; \left[h_{np}^{rs}(t) - h_{np}^{rs^*}(t)\right] dt \;\gtrsim\; 0; \quad \forall h_{np}^{rs}(t) \in \mathcal{H} \tag{4.39}$$

The lower level problem of the users' route choices is represented by the VI problem $\mathbf{VI}[E(g,h), h \in \mathcal{H}]$, which is defined in (P4). The solution of the lower level VI problem, $h^*(g)$, known as the reaction function, represents the users' best response for each given control g.

The above formulation (P9) is known as a Mathematical Program with Equilibrium Constraint (MPEC). The term "equilibrium constraint" refers to the variational inequality constraint of (4.39). MPEC is an extension of a bi-level program but it is more general. It represents a new class of optimization problems that have been applied extensively in engineering and economics. For readers interested in the MPEC method and its applications, see Luo et al. (1996).

The Stackelberg game of problem (P9) can also be formulated as the following *fixed point* model:

(P11)

$$g = F_g(g, h^*(g))$$

where,

$$h^*(g) \text{ solves } h = F_h(g, h)$$

Compared to the payoff in a Cournot equilibrium, the payoff for the traffic authority in a Stackelberg equilibrium must be higher (or at least equal) since the users' response has been taken into account when selecting the optimal control strategy.

4.4.3 System-Optimal Control-Assignment Problem: A Monopoly Game

The combined control-assignment problem is system-optimal if the DTA problem is system-optimal. Such a system-optimal control-assignment problem is essentially a monopoly game, in which the sole player—the traffic authority—controls both the signal settings and the traffic flows. Figure 4-4 shows the *extensive form* of this Monopoly game. This monopoly game is formulated by using path variables as follows:

(P12)

$$\min_{g,h} \quad Z_g = \int_0^T \sum_n \sum_{r,s} \sum_p c_p^{rs}(t)\; h_{np}^{rs}(t)\; dt \tag{4.40}$$

$$s.t. \;\; g_i^m(t) \in \mathcal{G}, \;\; h_{np}^{rs}(t) \in \mathcal{H} \tag{4.41}$$

The path cost $c_p^{rs}(t)$ depends on both flow vector, h, and control vector, g. The fixed point formulation of the monopoly game is given by:

(P13)

$$(g,h) = F_g(g,h)$$

where F_g is a mapping that gives the dynamic system-optimal (DSO) control and flow.

In order to derive the equivalent VI formulation, we first re-formulate the monopoly game of (P12) using *link variables* as follows:

(P14)

$$\min_{u,g} Z_g = \int_0^T \sum_{a \in \mathcal{A}} u_a(t) \tau_a[x_a(t), u_a(t), g_a(t)] dt \tag{4.42}$$

$$s.t. \;\; g_a(t) \in \mathcal{G}; \;\; u_a(t) \in \mathcal{U} \tag{4.43}$$

where $\mathcal{U}$ is the set of all feasible link flows and is defined by the dynamic network constraints of (3.17)–(3.26). The equivalent VI formulation of program (P14) is then given by:

(P15)

$$\int_0^T \sum_a u_a^*(t) \frac{\partial \tau_a^*(t)}{\partial g_a(t)} [g_a(t) - g_a^*(t)]\, dt +$$

$$\int_0^T \sum_a [\tau_a^*(t) + u_a^*(t) \frac{\partial \tau_a^*(t)}{\partial u_a(t)}] [u_a(t) - u_a^*(t)]\, dt \geq 0$$

$$\forall g_a(t) \in \mathcal{G}; \;\; u_a(t) \in \mathcal{U} \tag{4.44}$$

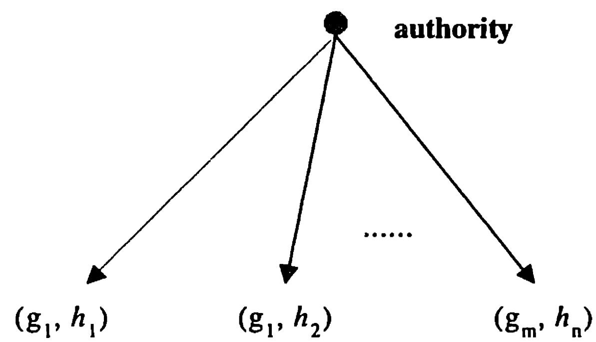

Figure 4-4: A Monopoly game for a single player—a traffic authority

where the asterisk indicates the optimal solution.

The outcome of the monopoly game represents the best performance that a system can ever achieve and thus serves as a benchmark for other solutions, such as Cournot and Stackelberg equilibria. In reality, however, a monopoly solution may not be sustained over the long run since it is not in the users' best interest. Therefore, the best possible strategy for a traffic authority is to seek a Stackelberg equilibrium which is superior to a Cournot equilibrium.

4.4.4 Other Cournot-Type Control Policies

All the above three models, one-level Cournot, bi-level Stackelberg, and system-optimal Monopoly are based on the assumption that the control objective is to minimize total travel delay (or total travel time). Such a control policy is known as *delay minimization* control. However, the models formulated above are not limited to delay minimization control and can be easily modified to accommodate other control policies. As reviewed in Chapter 2, there are two other popular control policies besides delay minimization, namely *Webster control* and *Smith's P_0 control.* Both control policies fall into the category of the one-level Cournot game when combined with the traffic assignment.

Webster Control

It is very easy to modify the one-level Cournot game formulation of Program (P7) to accommodate the Webster control. Webster control differs from delay minimization control in the definition of the *marginal link delay.* Recall that the objective of Webster control is to equalize the degree of saturation for each phase in an intersection. Therefore, the marginal link delay $\tilde{\tau}_a(t)$ under Webster control is defined as follows:

$$\tilde{\tau}_a(t) = -\frac{u_a(t)}{s_a(t)g_a(t)} \tag{4.45}$$

where $s_a(t)$ is the saturation flow of link a at time t. The marginal phase delay $\tilde{c}_i^m(t)$ is then given by:

$$\tilde{c}_i^m(t) = \sum_a \theta_{ia}^m(t)\, \tilde{\tau}_a(t) = \sum_a \theta_{ia}^m(t)\, \frac{u_a(t)}{s_a(t)g_a(t)} \tag{4.46}$$

By substituting the above $\tilde{c}_i^m(t)$ definition into VI problem of (P7), we have the formulation of the one-level Cournot game under Webster control. Interestingly, the objective of the Webster control can be viewed as miminmizing the following artificial term:

$$Z_{Webster} = \int_0^T \left\{ \int \tilde{\tau}_a(t)\; dg_a(t) \right\} dt \tag{4.47}$$

$$= \int_0^T \left\{ \int -\frac{u_a(t)}{s_a(t)g_a(t)}\; dg_a(t) \right\} dt \tag{4.48}$$

$$= -\int_0^T \frac{u_a(t)}{s_a(t)} \ln g_a(t)\, dt \tag{4.49}$$

where $\ln g_a(t) < 0$ since $0 \leq g_a(t) \leq 1$. Therefore, the above objective $Z_{Webster}$ is non-negative.

Smith's P_0 Control

Similarly, we can modify the VI problem (P7) to accommodate Smith's P_0 control. The objective of Smith's P_0 control, as defined by Smith (1981a), is to equalize the *cost of the clique* for each stage. Later Smith and Ghali (1990) redefined the term as *stage pressure.* The concept of a *clique* or a *stage* is identical to the concept of a *phase* used in the U.S. The link pressure is defined as the product of the saturation flow and the link travel time:

$$\tilde{\tau}_a(t) = -s_a(t)\tau_a(t) \tag{4.50}$$

The stage pressure is then given by:

$$\tilde{c}_i^m(t) = \sum_a \theta_{ia}^m(t)\, \tilde{\tau}_a(t) = -\sum_a \theta_{ia}^m(t)\, s_a(t)\tau_a(t) \tag{4.51}$$

The above stage pressure defines the marginal phase delay in the dynamic traffic control problem under Smith's P_0 control. Similarly, Smith's P_0 control can be viewed as minimizing the following artificial objective function:

$$Z_{Webster} = \int_0^T \left\{ \int \tilde{\tau}_a(t)\, dg_a(t) \right\} dt \tag{4.52}$$

$$= -\int_0^T s_a(t) \left\{ \int \tau_a(t)\, dg_a(t) \right\} dt \tag{4.53}$$

The combined P_0 control and dynamic traffic assignment problem is also a one-level Cournot game and its VI formulation is given by Program (P7) by substituting the marginal phase delay $c_i^m(t)$ with the one defined in Equation 4.51.

4.5 Summary

In this chapter, we propose a game theoretic methodology to integrate the dynamic traffic control (DTC) and the dynamic traffic assignment (DTA) problem. We formulate the DTC and DTA problems individually and analyze their optimality conditions. Then we present three sets of models of the combined dynamic traffic control-assignment (DTCA) problem. These three models are based on Cournot, Stackelberg and Monopoly games between a traffic authority and highway users.

A Stackelberg equilibrium solution is superior to that of a Cournot because the traffic authority anticipates users' reactions to the future control settings when solving the dynamic traffic control problem. The outcome of the monopoly game represents the optimal system performance and thus serves as a benchmark for other solutions, such as the Cournot and the Stackelberg equilibria.

The DTCA problem in this chapter is first formulated based on the delay minimization control objective. Other control policies, including Webster control and

Smith's P_0 control, can be combined with traffic assignment as one-level Cournot game models.

Chapter 5

Solution Algorithms for the Combined Dynamic Traffic Control-Assignment Models

In the previous chapter, we formulated a set of models for the combined dynamic traffic control–assignment (DTCA) problem. The objective of this chapter is to develop solution algorithms to solve those DTCA models.

5.1 Dynamic Traffic Control Algorithm

The dynamic traffic control (DTC) problem is to optimize the time-dependent green time splits given the cycle times, phases and link traffic flows. The continuous-time VI formulation of the DTC problem is given in Program (P3) in Section 4.2.1. It can be discretized as follows:

(P19)

$$\sum_{k=0}^{K}\sum_{i}\sum_{m}\tilde{c}_i^{m^*}(k)(g_i^m(k) - g_i^{m^*}(k)) \geq 0; \qquad \forall g_i^m(k) \in \mathcal{G} \tag{5.1}$$

The feasible set $\mathcal{G}$ is defined by the following constraints:

$$\sum_m g_i^m(k) = 1, \qquad \forall i, m \tag{5.2}$$

$$g_a(k) = \sum_m g_i^m(k)\, \theta_{ia}^m(k), \qquad \forall i, a \in B(i), m \tag{5.3}$$

$$g_i^m(k) \geq 0, \qquad \forall i, m \tag{5.4}$$

A different control policy defines a different marginal phase delay $\tilde{c}_i^{m^*}(k)$. The policy of delay minimization defines the marginal phase delay as:

$$\begin{aligned} \tilde{c}_i^m(k) &= \sum_a \theta_{ia}^m(k)\, \tilde{\tau}_a(k) \\ &= \sum_a \theta_{ia}^m(k)\, u_a(k)\, \frac{\partial \tau_a(k)}{\partial g_a(k)} \end{aligned} \tag{5.5}$$

The marginal phase delay under Webster control is given by:

$$\begin{aligned} \tilde{c}_i^m(k) &= \sum_a \theta_{ia}^m(k)\, \tilde{\tau}_a(k) \\ &= \sum_a \theta_{ia}^m(k)\, \frac{u_a(k)}{s_a(k) g_a(k)} \end{aligned} \tag{5.6}$$

and under Smith's P_0 control it is given by:

$$\begin{aligned} \tilde{c}_i^m(k) &= \sum_a \theta_{ia}^m(k)\, \tilde{\tau}_a(k) \\ &= \sum_a \theta_{ia}^m(k)\, s_a(k) \tau_a(k) \end{aligned} \tag{5.7}$$

The marginal phase delay for the Stackelberg model is more complicated since it depends on both the control and flow variables. It will be derived later in this chapter.

Recall that the dynamic system optimum (DSO) control derived in Section 4.2.2 is: *for each intersection, green time can only be assigned to a phase with the minimal marginal delay.* This is similar to the deterministic dynamic user-optimum (DUO) principle which states that *for each O-D pair, flow can only be assigned to a path with the minimal path travel cost.* Analogous to the famous *all-or-nothing* assignment in the deterministic traffic assignment procedure, the dynamic system optimum (DSO) control can be obtained by an *all-or-nothing* control. That is, at each iteration, for each intersection, the green time split for the phase with the minimal marginal delay is set to one and the green splits for other phases are set to zero. In addition, a smoothing between the current solution and the new solution is needed in order to prevent the swinging effect. A convex combination smoothing is adopted to guarantee the feasibility of the solution at each iteration. The DTC algorithm is outlined below.

Algorithm 1 **DTC Algorithm**

Step 1: *Initialize:*

1.1: Distribute equal green splits:

$g_i^{m(0)}(k) = 1/M, \; m = 1, \dots, M;$

M *= number of phases at junction* i*;*

1.2: Set counter $n = 0$.

Step 2: *Main loop:*

2.1:Compute the marginal phase delay $\tilde{c}_i^m(k)$ *based on green split solution* $g_i^{m(n)}(k)$*;*

2.2: Find the phase q *with the minimal marginal phase delay:*

$\tilde{c}_q^m(k) = \pi_i(k) = \min_{m \in \mathcal{M}_i} \tilde{c}_i^m(k)$

2.3: Perform an All-Or-Nothing control for each intersection i*:*

$g_i^{m(new)}(k) = 1;$ *for* $m = q$*;*

$g_i^{m(new)}(k) = 0;$ *for all* $m \neq q$*;*

2.4: Update green splits:

$g_i^{m(n+1)}(k) = g_i^{m(n)}(k) + \alpha^{(n)}(g_i^{m(new)}(k) - g_i^{m(n)}(k));$

Step 3: *Check convergence:*

If $g_i^{m(n+1)}(k) \approx g_i^{m(n)}(k)$, *then stop;*

Otherwise, set $n = n + 1$ *and go to Step 2.*

The step size α^n at Step 2.4 must be between zero and one and it can be calculated by either the simple formula of $\alpha^n = \frac{1}{n+1}$ or a more sophisticated line search algorithm such as that used in Frank-Wolfe algorithm (Frank and Wolfe 1956).

5.2 Dynamic Traffic Assignment Algorithm

The dynamic traffic Assignment (DTA) problem is to find the time-dependent traffic flows that are consistent with the behavior of multi-class users, given the time-dependent O-D demand and the control settings. We use the algorithm developed by Chabini and He (1998a, 1998b) to solve the DTA problem. The DTA algorithm consists of the C-Logit Route Choice algorithm (Section 5.2.1), the DNL algorithm (Section 5.2.2), and a convex combination smoothing. The DTA algorithm is presented below.

Algorithm 2 **DTA Algorithm**

Step 1: *Initialize*

J =maximum number of iterations;

Assign initial path flows $\{h_{np}^{rs^{(0)}}(k)\}$ based on free-flow path travel times;

$j = 0$.

Step 2: *Main loop*

2.1: Perform dynamic network loading using the DNL algorithm

2.2: Compute route choice probability $p_{np}^{rs}(k)$ for each user class by the C-Logit Route Choice algorithm;

2.3: Determine step size:

$\alpha^{(j)} = \frac{1}{j+1}$

2.3: Update path flows:

$h_{np}^{rs^{(j+1)}}(k) = h_{np}^{rs^{(j)}}(k) + \alpha^{(j)}[g_{np}^{rs}(k) - h_{np}^{rs^{(j)}}(k)],$

Step 3: *Check convergence:*

If $h^{(j+1)}(k) \approx h^{(j)}(k)$ *or* $j = J$*, then stop;*

Otherwise, $j = j + 1$ *and go to Step 2.*

where the DNL algorithm in Step 2.1 and the C-Logit Route Choice algorithm in Step 2.2 are explained in detail next.

5.2.1 C-Logit Route Choice Algorithm

A route choice procedure is needed to solve the user behavior model (see Figure 3-2). The objective of this model is to compute path choice probabilities for each O-D pair and each class of users, given the dynamic O-D demands and the path-based network conditions.

As proposed in Section 3.3.2, there are four classes of users: habitual, unguided, imperfectly guided and perfectly guided users. As discussed in Chapter 3, the deterministic dynamic user-optimal (DUO) route choice condition for class-4 users is essentially a special case of the stochastic dynamic user-optimal (SDUO) route choice condition for class-2 and class-3 users, in the sense that class-4 users have *zero* perception errors. Therefore, a common procedure to calculate the path choice probabilities can be applied to the last three classes of users. For class-1 users, the solution is trivial since their paths are pre-determined.

We denote $P_{np}^{rs}(k)$ as the class-n users' probability of choosing path p of O-D pair $(r,\ s)$ at time interval k and $h_{np}^{rs}(k)$ is the resulting flow on that path. That is,

$$h_{np}^{rs}(k) = D_n^{rs}(k)\, P_{np}^{rs}(k), \quad \forall r, s, p \tag{5.8}$$

where $D_n^{rs}(k)$ is the O-D demand of class-n users.

The probability $P_{np}^{rs}(k)$ is calculated based on the following C-Logit model (see Section 3.3.2 of Chapter 3):

$$P_{np}^{rs}(k) = \frac{exp[-\alpha_n c_p^{rs}(k) - CF_p]}{\sum_q exp[-\alpha_n c_q^{rs}(k) - CF_q]} \tag{5.9}$$

where CF_p is a *commonality factor* for path p and α_n is a positive parameter for class n users. On the one hand, the C-Logit model overcomes the Logit model's undesirable property of independence of irrelevant alternative (IIA), which is unacceptable for a route choice problem with path overlapping. On the other hand, the C-Logit model retains the Logit model's closed analytical structure.

The commonality factor CF_p is calculated as follows:

$$CF_p = \beta \ ln \left[\sum_q \left(\frac{L_{pq}}{\sqrt{L_p L_q}} \right)^{\gamma} \right] \tag{5.10}$$

where L_p and L_q are the lengths of paths p and q respectively and L_{pq} is the length of the common links shared by paths p and q. β and γ are two positive parameters. The link lengths can be measured by physical link lengths, or better yet, the link-additive part of generalized cost.

Algorithm 3 C-Logit Route Choice Algorithm

Step 1: *Compute commonality factors:*

For each O-D pair (r,s) and each path $p \in \mathcal{P}^{rs}$ do:

$CF_p = \beta \ ln \left[\sum_q \left(\frac{L_{pq}}{\sqrt{L_p L_q}} \right)^{\gamma} \right]$

end do

Step 2: *Compute path travel costs*

for each O-D pair (r,s) and each path $p \in \mathcal{P}^{rs}$ do:

for each time interval k = 1:K do

$c_p^{rs}(k) = \tau_{a1}(k) + \tau_{a2}(k + \tau_{a1}(k)) + \tau_{a3}(k + \tau_{a1}(k) + \tau_{a2}(k + \tau_{a1}(k))) + \ldots$

end do

end do

Step 3: *Compute route choice probabilities and path flows:*

for each O-D pair (r,s) and each path $p \in \mathcal{P}^{rs}$ do:

for each user class n do:

for each time interval k = 1:K do

$P^{rs}_{np}(k) = \frac{exp[-\alpha_n c^{rs}_p(k) - CF_p]}{\sum_q exp[-\alpha_n c^{rs}_q(k) - CF_q]}$

$h^{rs}_{np}(k) = D^{rs}_n(k)\, P^{rs}_{np}(k)$

end do

end do

end do

5.2.2 Dynamic Network Loading Algorithm

As discussed in Section 3.3.3, the dynamic network loading (DNL) problem is to find time-dependent link flows and link travel costs, given the route choices. In other words, we must find link variables that satisfy the dynamic network flow constraints of (3.17)–(3.26).

The discrete dynamic network flow constraints based on Equations (3.17)–(3.26) are derived as follows:

Link dynamics constraints:

$$x^{rs}_{ap}(k+1) = x^{rs}_{ap}(k) + u^{rs}_{ap}(k) - v^{rs}_{ap}(k), \qquad \forall a, r, s \tag{5.11}$$

Flow conservation constraints:

$$u^{rs}_{ap}(k) = \begin{cases} h^{rs}_p(k), & a \in \mathcal{A}(r) \ (a \text{ is the first link on path } p) \\ v^{rs}_{a'p}(k), & a \notin \mathcal{A}(r) \text{ and } a \text{ is after } a' \end{cases} \tag{5.12}$$

Flow propagation constraints:

$$V^{rs}_{ap}(k) = \sum_{j \in \{j: 0 \le j\Delta + \tau_a(j) \le k\Delta\}} u^{rs}_{ap}(j)\ \Delta, \qquad \forall a, p, r, s \tag{5.13}$$

$$v^{rs}_{ap}(k) = \left[V^{rs}_{ap}(k) - V^{rs}_{ap}(k-1)\right]/\Delta, \qquad \forall a, p, r, s \tag{5.14}$$

$$U^{rs}_{ap}(k) = \sum_{j=0}^{k} u^{rs}_{ap}(j)\ \Delta, \qquad \forall a, p, r, s \tag{5.15}$$

Definitional constraints:

$$u_a(k) = \sum_{r,s,p} u_{ap}^{rs}(k), \qquad \forall a \qquad (5.16)$$

$$v_a(k) = \sum_{r,s,p} v_{ap}^{rs}(k), \qquad \forall a \qquad (5.17)$$

$$x_a(k) = \sum_{r,s,p} x_{ap}^{rs}(k), \qquad \forall a \qquad (5.18)$$

Boundary conditions:

$$x_{ap}^{rs}(0) = 0, \qquad \forall a, p, r, s \qquad (5.19)$$

$$U_{ap}^{rs}(0) = 0, \qquad \forall a, p, r, s \qquad (5.20)$$

$$V_{ap}^{rs}(0) = 0, \qquad \forall a, p, r, s \qquad (5.21)$$

Non-negativity conditions:

$$x_{ap}^{rs}(k), u_{ap}^{rs}(k), v_{ap}^{rs}(k) \geq 0, \qquad \forall a, p, r, s \qquad (5.22)$$

We use the *I-Load*[1] algorithm developed by Chabini and He (1998a, 1998b) to solve the DNL problem. At each iteration, the link travel time, $\tau_a(k)$, is temporarily fixed. Note that once $\tau_a(k)$ and path flows $h_p^{rs}(k)$ are given, there are only five sets of unknowns, $x_{ap}^{rs}(k)$, $u_{ap}^{rs}(k)$, $v_{ap}^{rs}(k)$, $U_{ap}^{rs}(k)$, and $V_{ap}^{rs}(k)$ and five sets of equations, (5.11)-(5.15), in the DNL problem. Therefore a unique solution exists. The I-Load DNL algorithm is outlined as follows:

Algorithm 4 **DNL Algorithm**

Step 1: *Initialize:*

N = *maximum number of iterations;*

$\{\tau_a^{(0)}(k)\}$ = *free flow travel time;*

[1] I-load is named for iterative dynamic network loading

$n = 0.$

Step 2: *Main loop:*

2.1: Fix link travel time: $\tau_a(k) = \tau_a^{(n)}(k)$

2.2: Load path flows:

for each O-D (r, s) *do:*

for each path p *of O-D* (r, s) *do:*

for each sorted link a *on path* p *do:*

Compute $u_{ap}^{rs}(k)$ *using (5.12);*

Compute $U_{ap}^{rs}(k)$ *using (5.15);*

Compute $V_{ap}^{rs}(k)$ *using (5.13);*

Compute $v_{ap}^{rs}(k)$ *using (5.14).*

Compute $x_{ap}^{rs}(k)$ *using (5.11);*

end do

end do

end do

2.3 Aggregate link flows:

Compute $u_a(k)$ *using (5.16);*

Compute $v_a(k)$ *using (5.17);*

Compute $x_a(k)$ *using (5.18);*

2.4: Compute new link travel times $\{\tau_a^{new}(k)\}$*;*

2.5: Update link travel times:

$\tau_a^{(n+1)}(k) = \tau_a^{(n)}(k) + \alpha^{(n)}(\tau_a^{new}(k) - \tau_a^{(n)}(k));$

$\alpha^{(n)} = \frac{1}{n+1}.$

Step 3: *Check stopping criterion*

If $\tau_a^{(n+1)}(k) \approx \tau_a^{(n)}(k)$ *or* $n = N$, *then stop;*

Otherwise, $n = n + 1$ *and go to Step 2.*

It should be noted that the loading of departure flow along path p in Step 1.1 is done through sorted links along the path, starting from the link that connecting from origin r and ending at the link connecting to destination d. It is assumed that there is no cycle on any path under loading. This is not a restricted assumption since any path with a cycle is longer than the same path without the cycle and therefore a path with cycles can always be modified to a path without any cycle.

5.3 Solution Algorithms for the Combined Control-Assignment Models

5.3.1 Discrete Cournot Model

Using discrete time, the one-level DTCA model of the Cournot game [Program (P7)] is converted into the following VI problem:

(P16)

$$\sum_{k=0}^{K}\sum_{i,m}\tilde{c}_i^{m*}(k)[g_i^m(k)-g_i^{m*}(k)]+\sum_{0}^{T}\sum_{n,p,r,s}c_p^{rs*}(k)\left[h_{np}^{rs}(k)-h_{np}^{rs*}(k)\right]\geq 0$$

$$\forall g_i^m(k)\in\mathcal{G},\ h_{np}^{rs}(k)\in\mathcal{H} \tag{5.23}$$

where $\tilde{c}_i^{m*}(k)$ is the *marginal phase delay*, the derivative of the DTC objective function Z_g (Equation 4.9) with respective to the phase green time $g_i^m(k)$ at time k. In order to compare this with the gradient of the Stackelberg model, we re-define the objective function Z_g using path variables as follows:

$$\min_{g}\ \ Z_g=\sum_{k=0}^{K}\sum_{n}\sum_{r,s}\sum_{p}c_p^{rs}(k)\,h_{np}^{rs}(k) \tag{5.24}$$

The derivative of the objective function Z_g with respective to the phase green time

$g_i^m(l)$ at time l is given by:

$$\tilde{c}_i^m(l) = \frac{\partial Z_g}{\partial g_i^m(l)} \tag{5.25}$$

$$= \sum_{a \in B(i)} \frac{\partial Z_g}{\partial g_a(l)} \frac{\partial g_a(l)}{\partial g_i^m(l)} \tag{5.26}$$

$$= \sum_{a \in B(i)} \frac{\partial Z_g}{\partial g_a(l)} \theta_{ia}^m(l) \tag{5.27}$$

where the summation is over all the incoming links of node i. Equation (5.27) is obtained from the relationship between the link green split $g_a(l)$ and the phase green split $g_i^m(l)$ (Equation 4.11).

From (5.24), we have

$$\frac{\partial Z_g}{\partial g_a(l)} = \sum_{k=0}^{K} \sum_{n} \sum_{r,s} \sum_{p} h_{np}^{rs}(k) \frac{\partial c_p^{rs}(k)}{\partial g_a(l)} \tag{5.28}$$

because the path flow $h_{np}^{rs}(k)$ is given for the Cournot model. Furthermore,

$$\frac{\partial c_p^{rs}(k)}{\partial g_a(l)} = \begin{cases} \frac{\partial c_p^{rs}(k)}{\partial \tau_a(l)} \frac{\partial \tau_a(l)}{\partial g_a(l)}, & \text{if } k + c_p^{ra}(k) = l \\ 0, & \text{otherwise} \end{cases} \tag{5.29}$$

$$\frac{\partial c_p^{rs}(k)}{\partial \tau_a(l)} = \begin{cases} 1, & \text{if } k + c_p^{ra}(k) = l \\ 0, & \text{otherwise} \end{cases} \tag{5.30}$$

where $c_p^{ra}(k)$ is the path traversal time from origin r to the tail node of link a along path p departing at time k. It can be calculated recursively using the formula given in Equation (3.34). Equations (5.28)–(5.30) assume that the First-In-First-Out (FIFO) condition[2] is satisfied. Combine (5.29) and (5.30), we have

$$\frac{\partial c_p^{rs}(k)}{\partial g_a(l)} = \begin{cases} \frac{\partial \tau_a(l)}{\partial g_a(l)}, & \text{if } k + c_p^{ra}(k) = l \\ 0, & \text{otherwise} \end{cases} \tag{5.31}$$

Substitute (5.31) into (5.28), we have

$$\frac{\partial Z_g}{\partial g_a(l)} = \sum_n \sum_{r,s} \sum_p h_{np}^{rs}(k) \frac{\partial c_p^{rs}(k)}{\partial g_a(l)} = \sum_n \sum_{r,s} \sum_p h_{np}^{rs}(k) \frac{\partial \tau_a(l)}{\partial g_a(l)} \tag{5.32}$$

Thus, the marginal phase delay (gradient) in (5.25) is given by

$$\tilde{c}_i^m(l) = \frac{\partial Z_g}{\partial g_i^m(l)} \tag{5.33}$$

$$= \sum_{a \in B(i)} \theta_{ia}^m(l) \frac{\partial Z_g}{\partial g_a(l)} \tag{5.34}$$

$$= \sum_{a \in B(i)} \theta_{ia}^m(l) \sum_n \sum_{r,s} \sum_p h_{np}^{rs}(k) \frac{\partial \tau_a(l)}{\partial g_a(l)} \tag{5.35}$$

where $\frac{\partial \tau_a(l)}{\partial g_a(l)}$ can be obtained from the function form of τ_a.

The above results assume that the First-In-First-Out (FIFO) rule is satisfied. If the FIFO does not hold, the gradient result of Equation (5.35) becomes

$$\tilde{c}_i^m(l) = \sum_{a \in B(i)} \theta_{ia}^m(l) \sum_n \sum_{r,s} \sum_p \sum_{k: k + c_p^{ra}(k) = l} h_{np}^{rs}(k) \frac{\partial \tau_a(l)}{\partial g_a(l)} \tag{5.36}$$

5.3.2 Discrete Stackelberg Model

The discrete-time formulation of the bi-level DTCA model of the Stackelberg game [Program (P11)] is given by:

(P17)

[2]The First-In-First-Out (FIFO) rule is defined by

$$k + \tau(k) < k + \Delta k + \tau(k + \Delta k), \quad \text{if } \Delta k > 0$$

$$\min_{g(k)} \quad Z_g = \sum_{k=0}^{K} \sum_{n} \sum_{r,s} \sum_{p} c_p^{rs}(k)\, h_{np}^{rs^*}(k) \tag{5.37}$$

s.t.

$$g_i^m(k) \in \mathcal{G} \tag{5.38}$$

$$\sum_{k=0}^{K} \sum_{n} \sum_{r,s} \sum_{p} E_p^{rs^*}(k)\, [h_{np}^{rs}(k) - h_{np}^{rs^*}(k)] \geq 0; \quad \forall h_{np}^{rs}(k) \in \mathcal{H} \tag{5.39}$$

The derivative of the objective function Z_g with respective to the phase green time $g_i^m(l)$ at time l is given by:

$$\frac{\partial Z_g}{\partial g_i^m(l)} = \sum_{a \in B(i)} \frac{\partial Z_g}{\partial g_a(l)} \frac{\partial g_a(l)}{\partial g_i^m(l)} \tag{5.40}$$

$$= \sum_{a \in B(i)} \frac{\partial Z_g}{\partial g_a(l)} \theta_{ia}^m(l) \tag{5.41}$$

where the summation is over all the incoming links of node i. Equation (5.41) is obtained from the relationship between the link green split $g_a(l)$ and the phase green split $g_i^m(l)$ (Equation 4.11).

Furthermore,

$$\frac{\partial Z_g}{\partial g_a(l)} = \sum_{k=0}^{K} \sum_{n} \sum_{r,s} \sum_{p} \left[c_p^{rs}(k) \frac{\partial h_{np}^{rs}(k)}{\partial g_a(l)} + h_{np}^{rs}(k) \frac{\partial c_p^{rs}(k)}{\partial g_a(l)} \right] \tag{5.42}$$

$$\frac{\partial h_{np}^{rs}(k)}{\partial g_a(l)} = \sum_{q \in \mathcal{P}^{rs}} \frac{\partial h_{np}^{rs}(k)}{\partial c_q^{rs}(k)} \frac{\partial c_p^{rs}(k)}{\partial g_a(l)} \tag{5.43}$$

$$\frac{\partial c_p^{rs}(k)}{\partial g_a(l)} = \begin{cases} \frac{\partial c_p^{rs}(k)}{\partial \tau_a(l)} \frac{\partial \tau_a(l)}{\partial g_a(l)}, & \text{if } k + c_p^{ra}(k) = l \\ 0, & \text{otherwise} \end{cases} \tag{5.44}$$

$$\frac{\partial c_p^{rs}(k)}{\partial \tau_a(l)} = \begin{cases} 1, & \text{if } k + c_p^{ra}(k) = l \\ 0, & \text{otherwise} \end{cases} \tag{5.45}$$

and,

$$\frac{\partial c_p^{rs}(k)}{\partial g_a(l)} = \begin{cases} \frac{\partial \tau_a(l)}{\partial g_a(l)}, & \text{if } k + c_p^{ra}(k) = l \\ 0, & \text{otherwise} \end{cases} \tag{5.46}$$

From the definition of the path departure flow (Equation 4.27), we have

$$\frac{\partial h_{np}^{rs}(k)}{\partial c_q^{rs}(k)} = D^{rs}(k) \frac{\partial P_{np}^{rs}(k)}{\partial c_q^{rs}(k)} \tag{5.47}$$

Substitute (5.47) into (5.43), we have

$$\frac{\partial h_{np}^{rs}(k)}{\partial g_a(l)} = \begin{cases} D_n^{rs}(k) \sum_{q \in \mathcal{P}^{rs}} \frac{\partial P_{np}^{rs}(k)}{\partial c_q^{rs}(k)} \frac{\partial \tau_a(l)}{\partial g_a(l)}, & \text{if } k + c_p^{ra}(k) = l \\ 0, & \text{otherwise} \end{cases} \tag{5.48}$$

From (5.42), (5.46), and (5.48), the gradient becomes

$$\frac{\partial Z_g}{\partial g_i^m(l)} = \sum_{a \in B(i)} \theta_{ia}^m(l) \sum_n \sum_{r,s} \sum_p \left[c_p^{rs}(k) \frac{\partial h_{np}^{rs}(k)}{\partial g_a(l)} + h_{np}^{rs}(k) \frac{\partial c_p^{rs}(k)}{\partial g_a(l)} \right]$$

$$= \sum_{a \in B(i)} \theta_{ia}^m(l) \sum_n \sum_{r,s} \sum_p \left[c_p^{rs}(k) D^{rs}(k) \sum_{q \in \mathcal{P}^{rs}} \frac{\partial P_{np}^{rs}(k)}{\partial c_q^{rs}(k)} \frac{\partial \tau_a(l)}{\partial g_a(l)} + h_{np}^{rs}(k) \frac{\partial \tau_a(k)}{\partial g_a(l)} \right]$$

$$= \sum_{a \in B(i)} \theta_{ia}^m(l) \frac{\partial \tau_a(l)}{\partial g_a(l)} \sum_n \sum_{r,s} \sum_p \left[c_p^{rs}(k) D^{rs}(k) \sum_{q \in \mathcal{P}^{rs}} \frac{\partial P_{np}^{rs}(k)}{\partial c_q^{rs}(k)} + h_{np}^{rs}(k) \right] \tag{5.49}$$

In the above equations, k satisfies $k + c_p^{ra}(k) = l$. Again, the gradient result of (5.49) assumes FIFO. If FIFO is not satisfied, Equation (5.49) becomes

$$\frac{\partial Z_g}{\partial g_i^m(l)} = \sum_{a \in B(i)} \theta_{ia}^m(l) \frac{\partial \tau_a(l)}{\partial g_a(l)} \sum_n \sum_{r,s} \sum_p \sum_{k:k+c_p^{ra}(k)=l} \left[c_p^{rs}(k) D^{rs}(k) \sum_{q \in \mathcal{P}^{rs}} \frac{\partial P_{np}^{rs}(k)}{\partial c_q^{rs}(k)} + h_{np}^{rs}(k) \right] \tag{5.50}$$

The above results are valid for any route choice model. For a C-Logit model, we can have some further results. Under the C-Logit model, the choice probability of path p, $P_{np}^{rs}(k)$ for class n users is given by Equation (5.9). Its derivative with respect to the actual path travel time $c_q^{rs}(k)$ on path q of the same O-D pair is given by

$$\frac{\partial P_{np}^{rs}(k)}{\partial c_q^{rs}(k)} = \frac{\partial \ln P_{np}^{rs}(k)}{\partial \ln c_q^{rs}(k)} \frac{P_{np}^{rs}(k)}{c_q^{rs}(k)}$$
$$= \begin{cases} -\alpha_n P_{np}^{rs}(k)\,[1 - P_{np}^{rs}(k)], & \text{for } p = q \\ \alpha_n P_{np}^{rs}(k) P_{nq}^{rs}(k), & \text{for } p \neq q \end{cases} \tag{5.51}$$

where α is the parameter for the path travel time in the C-Logit function.

5.3.3 Discrete Monopoly Model

Similarly, the discrete-time formulation of the System-Optimal DTCA model of the Monopoly game [Program (P12)] with path variables is given by:

(P18)

$$\min_{g,h} \quad Z_g(g,h) = \sum_{k=0}^{K} \sum_{n} \sum_{r,s} \sum_{p} c_p^{rs}(k)\; h_{np}^{rs}(k) \tag{5.52}$$

$$s.t. \quad g \in \mathcal{G}, \;\; h \in \mathcal{H} \tag{5.53}$$

There are two types of decision variables in the Monopoly model: green time split g, and path flow h. The derivative of $Z_g(g,h)$ with respective to green time split $g_i^m(l)$ is the same as the one in the Cournot model:

$$\frac{\partial Z_g}{\partial g_i^m(l)} = \sum_{a \in B(i)} \theta_{ia}^m(l) \sum_{n} \sum_{r,s} \sum_{p} h_{np}^{rs}(k) \frac{\partial \tau_a(l)}{\partial g_a(l)} \tag{5.54}$$

where k satisfies $k + c_p^{ra}(k) = l$. The difference between the Cournot and the Monopoly models is on the route choices. The Cournot model assigns flows according to the user-optimal conditions while the Monopoly model assigns flows based on the system-optimal conditions.

5.3.4 DTCA Solution Algorithms

In order to solve each of the above combined dynamic traffic control-assignment (DTCA) models, we decompose the DTCA problem into two subproblems: DTC and DTA. These two subproblems can be solved by the DTC algorithm in Section 5.1 and the DTA algorithm in Section 5.2.

We use a gradient-based method to solve the DTCA models (see Bertsekas 1995, Chen 1993 and Luo et al. 1996 for a review on nonlinear and bi-level programming algorithms). We first present an exact algorithm and then give two simplified approximations.

The exact algorithm (Algorithm DTCA-1) for solving the DTCA problem is given as follows:

Algorithm 5 Algorithm DTCA-1

Step 1: *Initialize:*

1.1: Distribute equal green splits:

$g_i^{m(0)}(k) = 1/M, \ m = 1, \dots, M;$

M=number of phases at junction i.

1.2: Set counter $I = 0$.

Step 2: *All-Or-Nothing Control (Gradient Calculation):*

2.1: Solve the DTA subproblem for $g = g^{(I)}$ *using the DTA algorithm; yield a DTA solution* $h^{(I)}, x^{(I)}, u^{(I)}$.

2.2: Compute the marginal phase delay (gradient) $\tilde{c}_i^m(l) = \frac{\partial Z_g}{\partial g_i^m(l)}$ *at* $h = h^{(I)}$, $x = x^{(I)}$, $u = u^{(I)}$, *and* $g = g^{(I)}$ *using the following formula:*

Equation 5.32 for the Cournot model;

Equation 5.49 for the Stackelberg model; and

Equation 5.54 for the Monopoly model.

2.2: Find the phase q with the minimal marginal phase delay:

$\tilde{c}_q^m(k) = \pi_i(k) = \min_{m \in \mathcal{M}_i} \tilde{c}_i^m(k)$.

2.3: Perform an All-Or-Nothing *control for each intersection i:*

$\hat{g}_i^m(k) = 1;$ *for* $m = q;$

$\hat{g}_i^m(k) = 0$; *for all* $m \neq q$.

Step 3: *Step Size Determination:*

Calculate the step size α^I *using the method of successive average (MSA) or Armijo rule.*

Step 4: *Update green splits:*

$$g_i^{m(I+1)}(k) = g_i^{m(I)}(k) + \alpha^I(\hat{g}_i^m(k) - g_i^{m(I)}(k));$$

Step 5: *Check convergence:*

If $g_i^{m(I+1)}(k) \approx g_i^{m(I)}(k)$, *then stop;*

Otherwise, set $I = I + 1$ *and go to Step 2.*

Algorithm DTCA-1 is illustrated in Figure 5-1. The *All-Or-Nothing* control is the procedure to assign the maximum green time split to the phase with the minimum marginal phase delay (the gradient of the DTC objective function). The gradient calculations for the three combined models are given in Equations (5.32),(5.49), and 5.54 respectively. The DTA subproblem is embedded in the gradient calculation (Step 2.1). At each iteration, we need to solve a complete DTA subproblem, which is very costly. Therefore, the bottleneck of Algorithm DTCA-1 is in Step 2—gradient calculation.

In order to overcome the bottleneck of Algorithm DTCA-1, we use the DTA solution in the previous iteration as an approximation in Step 2.1. Since the gradient calculation is based on an approximation, an outer loop for smoothing both control and flow variables is needed. This two-looped algorithm is termed Algorithm DTCA-2 and given in Figure 5-2. In the inner loop, the two subproblems of DTC and DTA are solved in sequence. Until the convergence of the inner loop is satisfied, the solutions of DTC and DTA are combined in the outer loop. Algorithm DTCA-2 is outline below:

Algorithm 6 **Algorithm DTCA-2**

Step 1: *Initialize:*

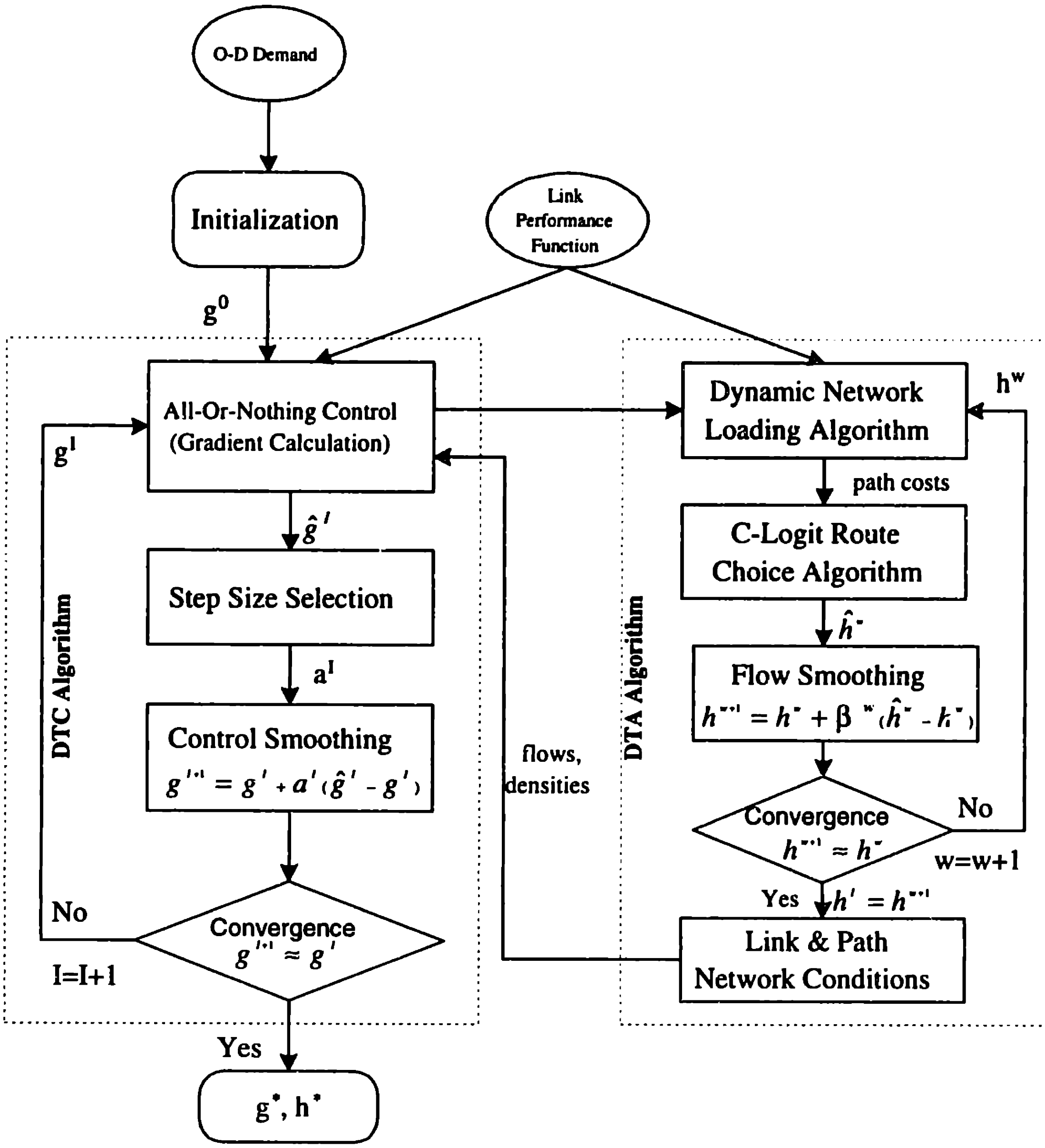

Figure 5-1: Algorithm DTCA-1

1.1: Distribute equal green splits:

$g_i^{m(0)}(k) = 1/M, \ m = 1, \ldots, M;$

M=number of phases at junction i.

1.2: Assign initial path flows $h_p^{rs(0)}(k)$ *based on free-flow travel times.*

1.3: Set counter $I = 0$.

Step 2: *Main Outer Loop:*

2.1: DTA inner loop:

Solve the DTA subproblem using the DTA algorithm;

yield a new DTA solution $\hat{h}_p^{rs(I)}(k)$.

2.2: DTC inner loop:

Solve the DTC subproblem using the DTC algorithm;

yield a new DTC solution $\hat{g}_i^{m(I)}(k)$.

2.3: Outer smoothing:

$$g_i^{m(I+1)}(k) = g_i^{m(I)}(k) + \gamma^I[\hat{g}_i^{m(I)}(k) - g_i^{m(I)}(k)];$$

$$h_p^{rs(I+1)}(k) = h_p^{rs(I)}(k) + \gamma^I[\hat{h}_p^{rs(I)}(k) - h_p^{rs(I)}(k)];$$

Step 3: *Outer convergence check:*

If $g_i^{m(I+1)}(k) \approx g_i^{m(I)}(k)$ *and* $h_p^{rs(I+1)}(k) \approx h_p^{rs(I)}(k)$, *then stop;*

Otherwise, set $I = I + 1$ *and go to Step 2.*

To further streamline the algorithm, we can conduct just one-iteration DTA for the gradient calculation step (Step 2.1 of Algorithm DTCA-1). The solutions of two subproblems are combined at the end of each iteration and the convergence is checked immediately. This streamlined approach is termed *Algorithm DTCA-3* and its diagram is depicted in Figure 5-3.

Algorithm 7 Algorithm DTCA-3

Step 1: *Initialize:*

1.1: Distribute equal green splits:

$g_i^{m(0)}(k) = 1/M, \ m = 1, \ldots, M;$

M=number of phases at junction i.

1.2: Assign initial path flows $h_p^{rs(0)}(k)$ based on free-flow travel times.

1.3: Set counter $I = 0$.

Step 2: *Solve the DTA subproblem in one iteration.*

Step 3: *Solve the DTC subproblem in one iteration.*

Step 3: *Convergence check:*

If $g_i^{m(I+1)}(k) \approx g_i^{m(I)}(k)$ and $h_p^{rs(I+1)}(k) \approx h_p^{rs(I)}(k)$, then stop;

Otherwise, set $I = I + 1$ and go to Step 2.

Notice that Algorithm DTCA-3 is a special case of Algorithm DTCA-2 when Algorithm DTCA-2's inner loop contains only one iteration.

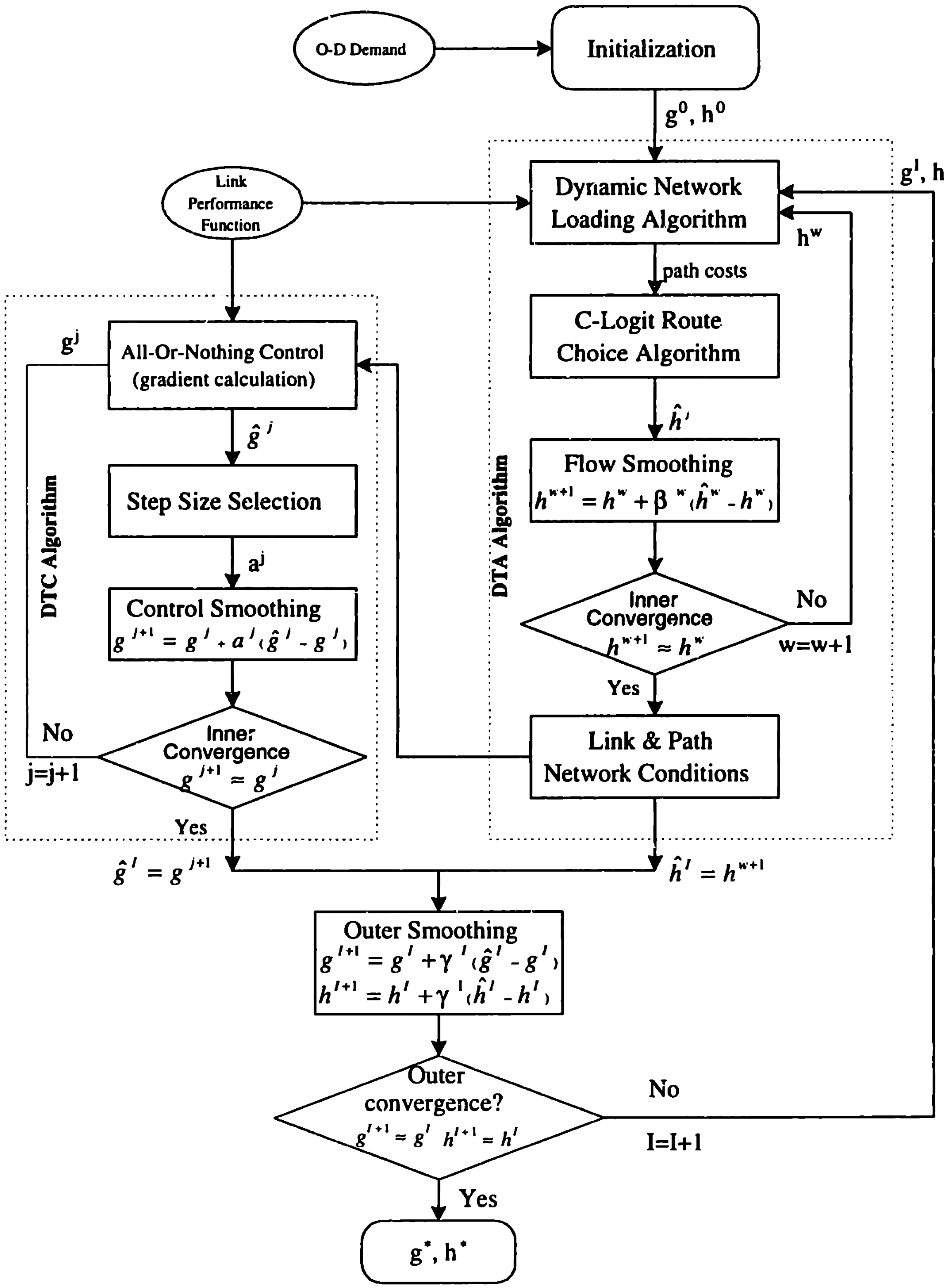

Figure 5-2: Algorithm DTCA-2

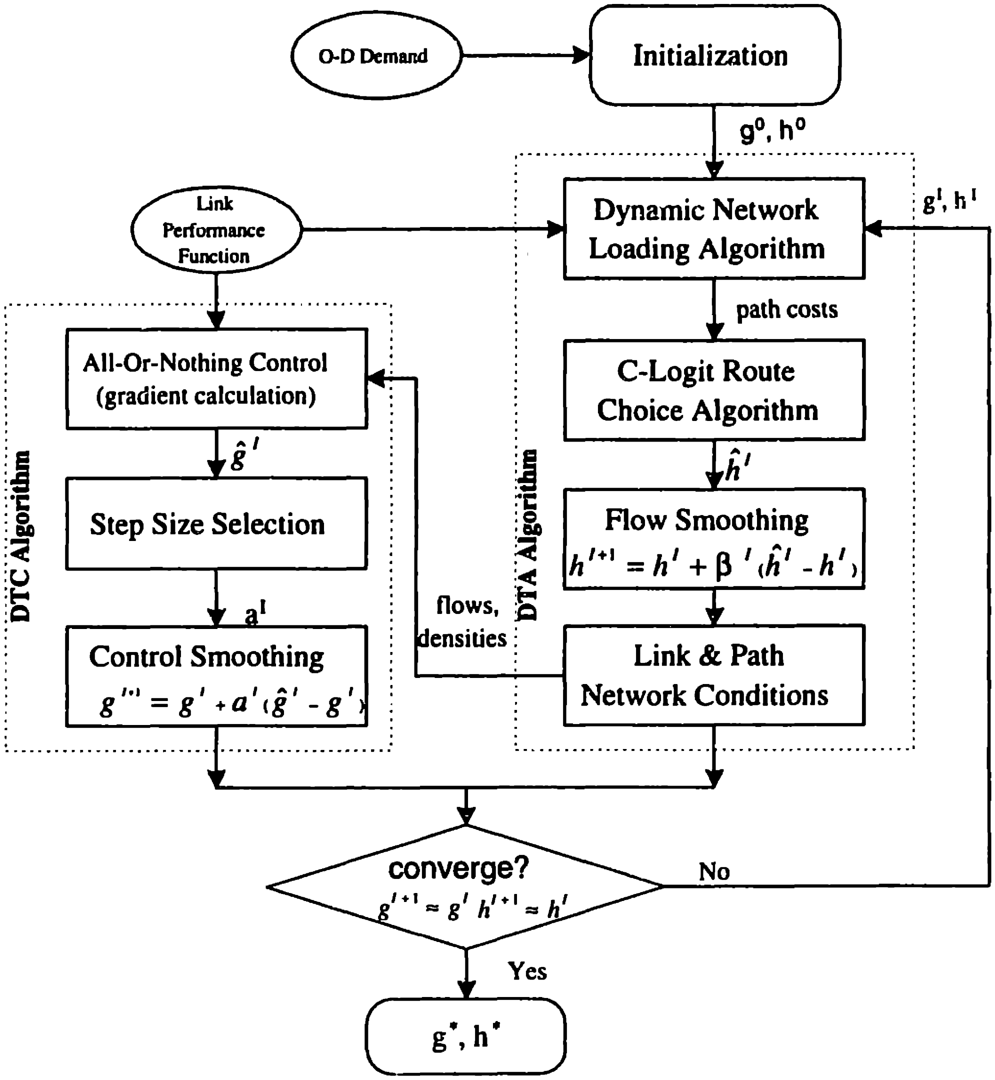

Figure 5-3: Algorithm DTCA-3

Chapter 6

A Case Study

In previous chapters, we developed a framework and a set of mathematical models to integrate dynamic traffic control and assignment. The objective of this chapter is to compare the performance of various models and evaluate the benefits of the combined control and assignment strategies in a real urban traffic network.

6.1 Case Study Design

The selected test network is a grid urban network from the Boston Back Bay. A map of the Boston Back Bay area is given in Figure 6-1.

The Back Bay network consists of four streets in an east-west direction and eight streets in an north-south direction. Figure 6-2 depicts the network under consideration. The network is modeled as a directed graph. Each turning movement represents a directed link. An intersection splits into two nodes, each with an outgoing direction. The detailed network representation is given in Figure 6-3. The network contains 130 links, 82 nodes, and 32 intersections, 26 of which are signalized.

There are 100 origin-destination (O-D) pairs (10 origins and 10 destinations) in the Back Bay network. The O-D demands used in this case study are estimated from the sensor count data for each turning movement. See Ashok (1996) for a review of O-D estimation methods. The following two statistics are computed to measure the errors of the O-D estimation:

1. Root Mean Squared Error (RMS)= $\sqrt{\frac{\sum_a (x_a - \hat{x}_a)^2}{N}}$

2. Root Mean Squared Normalized Error (RMSN) = $\frac{\sqrt{N \sum_a (x_a - \hat{x}_a)^2}}{\sum_a x_a}$

where x_a is the actual link volume for link a and $\hat{x}_a$ is the estimated volume implied by the estimated O-D demand. The summation is over all links. For the Back Bay network data, the above two statistics for the O-D estimation are:

- RMS = 55.549

- RMSN = 0.1435

These values indicate a good fit in the O-D estimation. Such small estimation errors are due to the detailed representation of the network. The O-D flows are calculated from the turning movement volumes rather than the usual link volumes.

The O-D data estimated from the turning movement volumes are static. In order to obtain multiple-time O-D demand, we MATLAB's random number generation function[*random(Normal, mean, std)*] to generation multiple O-D matrices with mean equal to the estimated static O-D demand and standard deviation equal to 5% of the mean. Each generated O-D matrix is used for a time period. Furthermore, the O-D demand is divided into three user classes: habitual (5%), unguided (85%) and guided (10%) users. The path travel times used in the C-Logit route choice model are measured in minutes. The scale parameter (α_n) in the C-Logit model is 1.0 for the unguided users and 3.0 for the guided users. The values of the parameters in the commonality factor calculation (Equaiton 5.10) are specified as: $\beta = 1.0$ and $\gamma = 2.0$.

Four paths are pre-determined as the choice set for each O-D pair. Therefore, there are total of 400 paths.

The time horizon under consideration is from 0 to 60 ($K = 60$). The time step used is 30 seconds ($\Delta = 30$ seconds).

This study tests two link performance functions. One is the linear queuing delay function as follows:

$$\tau_a(k) = \tau_{a0} + \frac{x_a(k)}{s_a\, g_a(k)} \tag{6.1}$$

where τ_{a0} is the free-flow travel time; s_a is the saturation flow rate; $x_a(k)$ is the number of vehicles on link a at time interval k; and $g_a(k)$ is the green time split for link a at time interval k.

The other link performance function used is the *Highway Capacity Manual* (HCM, 1994) delay function, defined as follows:

$$\tau_a(k) = 0.38\, C_i(k) \frac{[1 - g_a(k)]^2}{1 - \frac{u_a(k)}{s_a}} + 173\, (\rho_a(k))^2 \left[(\rho_a(k) - 1) + \sqrt{(\rho_a(k) - 1)^2 + 16 \frac{\rho_a(k)}{s_a}} \right] \tag{6.2}$$

where $C_i(k)$ is the cycle length of intersection i (to which link a belongs); $u_a(k)$ is the entry flow of link a at time interval k; and $\rho_a(k) = \frac{u_a(k)}{s_a g_a(k)}$ is the degree of saturation for link a at time interval k .

The following six control policies are tested in this case study:

- Current existing pre-timed control
- Responsive Webster equal-saturation control
- Responsive Smith P_0 control
- One-level delay minimization control (Cournot)
- Bi-level delay minimization control (Stackelberg)
- System-optimal control (Monopoly)

The combined control-assignment models under the above six control policies are solved using the DTCA algorithms developed in Chapter 5. All of the above combined control-assignment models are solved using the two-looped algorithm DTCA-2. Algorithm DTCA-2 contains an inner loop and an outer loop. The inner loop solves two subproblems of DTC and DTA independently while the outer loop combines the solutions of DTC and DTA. To speed the running time, the inner loop is solved for the maximum of 10 iterations. A smoothing method is also used in the outer loop. An optimal solution is found when both the green time splits and the path flows converge in the outer loop. As discussed in Chapter 5, algorithm DTCA-3 is a special case of algorithm DTCA-2 when DTCA-2's inner loop contains only one iteration. The algorithms are coded in MATLAB.

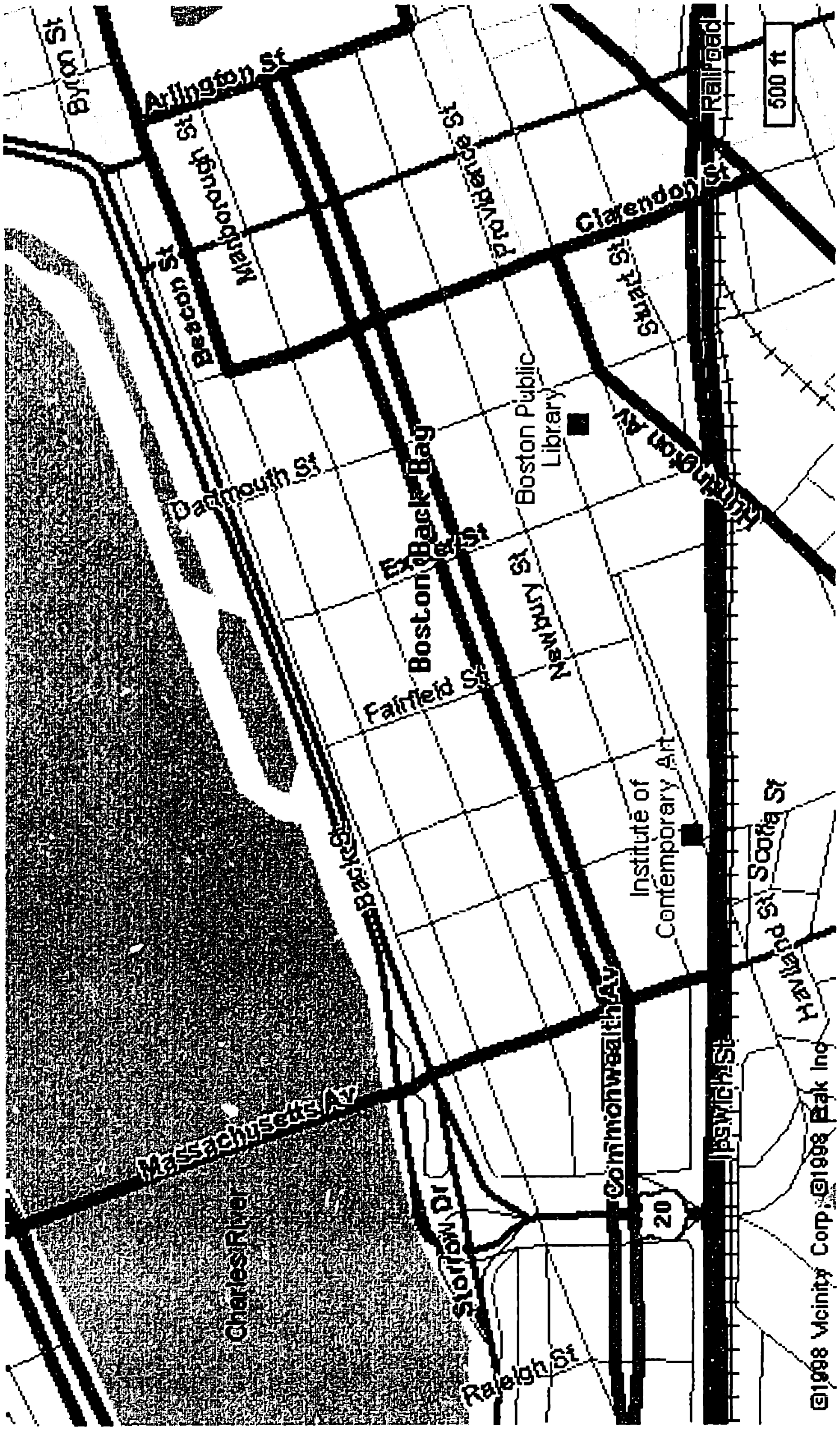

Figure 6-1: Boston Back Bay Area

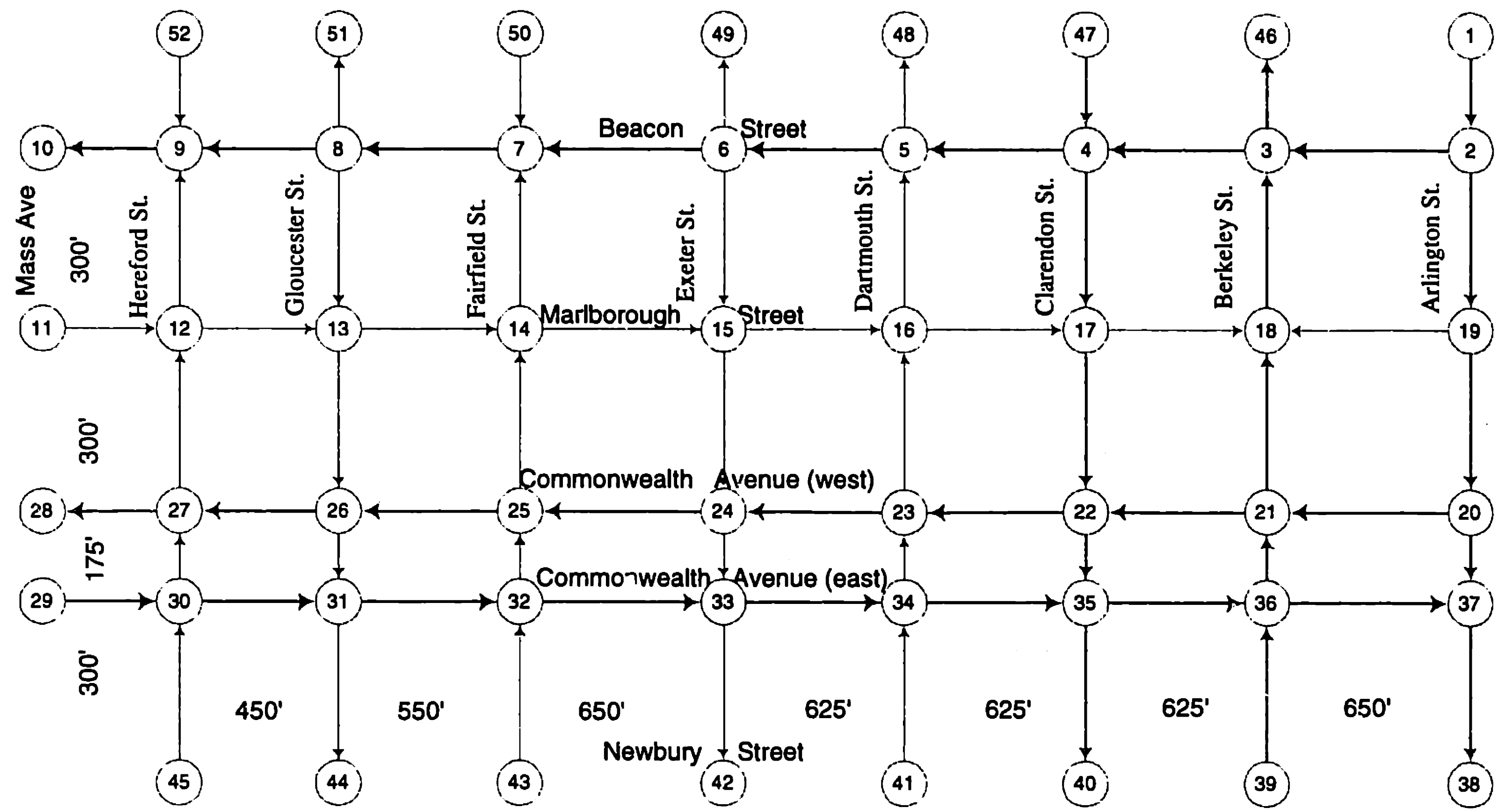

Figure 6-2: Boston Back Bay Network

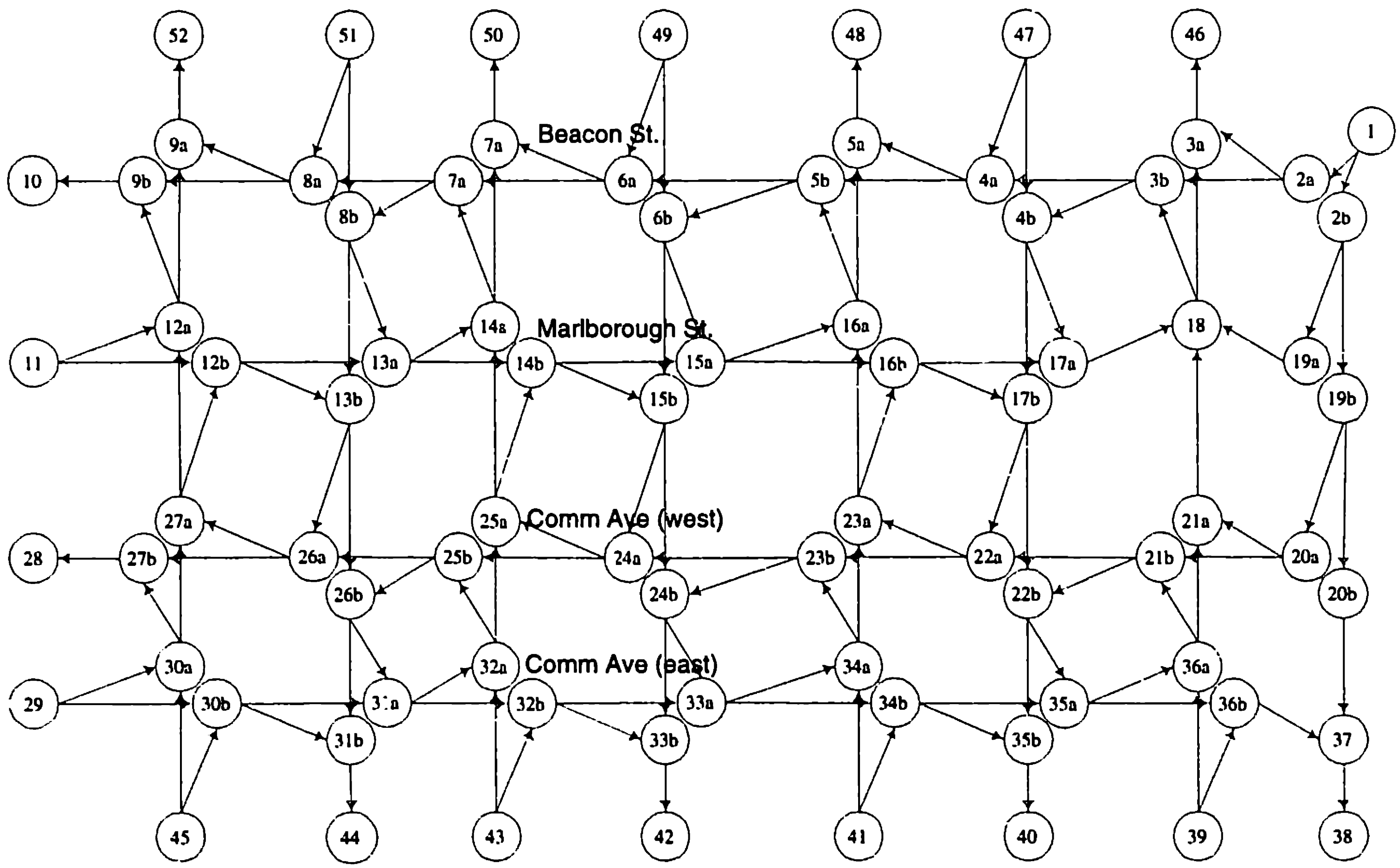

Figure 6-3: Detailed Network Representation

To validate the model's results, we compare link speeds given by the model under the existing pretimed control with the link speeds obtained from the sensors. Unfortunately, we do not have the sensor data for all links. Only 14 links have sensor data on speed. Nevertheless, the *Root Mean Squared Error* (RMS) and the *Root Mean Squared Normalized Error* (RMSN) for these 14 links are obtained as follows:

- RMS = 3.3631

- RMSN = 0.2803

The above statistics indicate that there are some small errors between the actual link speeds and those produced by the model. Two sources may contribute to those errors. One is the difference between the actual link flow and the estimated link flow. The other is that the link performance function may not accurately capture the actual link delays.

6.2 Results and Analysis

The following performance measures are calculated for each control strategy:

- Total travel time;

- Average link travel time;

- Average link speed; and

- Average path travel time.

The total travel time is the most important measurement in this case study because it is the objective function of the three combined control-assignment models. In addition to the above statistics, the green times and the corresponding link flows over the entire time horizon are also given at an intersection as an example.

6.2.1 Results with the Linear Queuing Delay Function

Table 6.1 presents the total travel time results of the six control policies when the link performance function is given by the linear queuing delay. The Monopoly solution is the system-optimal solution and it represents the best performance. The optimality gaps of other solutions to the monopoly solution are also shown in Table 6.1. All three combined model solutions—Cournot, Stackelberg and Monopoly—reduce the total travel time significantly from the existing pretimed control. The Cournot solution reduces the total travel time by about 11% from the existing solution. The Stackelberg solution further decreases the total travel time over the Cournot solution. A small gap (1.3%) between the Cournot and Stackelberg solutions exists. The gap between the Stackelberg and the Monopoly represents the difference between a system-optimal flow and a user-optimal flow.

Table 6.1: Total Travel Time Results with the Linear Queuing Function

Controls	Total Travel Time (mins)	Optimality Gap (%)
Existing	11784	14.12
Webster	11781	14.10
Smith P_0	11566	12.02
Cournot	10642	3.07
Stackelberg	10504	1.73
Monopoly	10325	0

Table 6.2 shows the average link travel time and the average link speed results. The percentage changes given in the table are measured against the solution under the existing pretimed control. There is a small improvement in the link traffic conditions under Webster control. However, both the average link travel time and link speed under Smith P_0 control are worse than those under the pretimed control. Smith

Table 6.2: Link Travel Times and Speeds with the Linear Queuing Function

Controls	Avg Link TT (secs)	Change (%)	Avg Link Speed (mph)	Change (%)
Existing	35.82	0.00	8.65	0.00
Webster	35.06	-2.11	8.67	0.20
Smith P_0	36.09	0.76	8.60	-0.56
Cournot	33.29	-7.05	9.15	5.73
Stackelberg	32.93	-8.06	9.24	6.82
Monopoly	32.16	-10.21	9.50	9.71

P_0 control allocates green time to equalize the product of the saturation flow and travel time: $s_a \tau_a$. Therefore, a link with a large saturation flow results in a small link travel time, while a link with a small saturation flow has a large travel time. That is, Smith P_0 control creates a link travel time disparity. By contrast, all three combined solutions reduce the average link travel time and increase the average link speed significantly.

The average path travel time results are presented in Table 6.3. Consistent with link results in Table 6.2, the Smith P_0 control has on average a longer path travel time than the existing pretimed control, though the difference is very small. The Webster control shows a small improvement in the path level. Again, the three combined policies have large reductions in the average path travel time.

Figure 6-4 shows the signal setting at intersection 22 as an example. It draws the green time split for the westbound of Commonwealth Avenue that intersects with Clarendon Street (link 52 of intersection 22). The current existing control is a fixed-timed control. The green time split is constant at 0.61 during the entire time horizon. The other five controls—Webster, Smith, Cournot, Stackelberg and Monopoly—have time-varying green splits. Since the westbound Commonwealth Avenue is a major

Table 6.3: Path Travel Times with the Linear Queuing Function

Controls	Avg Path TT (mins)	Change (%)
Existing	5.57	0.00
Webster	5.45	-2.07
Smith P_0	5.60	0.47
Cournot	5.19	-6.81
Stackelberg	5.14	-7.72
Monopoly	5.03	-9.76

street compared to the conflicting southbound Clarendon Street, the westbound obtains a large share of green time (greater than 50%). The Monopoly control allocates the largest green split (around 0.7).

Figure 6-5 shows the flow volumes of the two approaching links in this intersection and Figure 6-6 depicts the travel times of the two corresponding links. Both the flow and the travel time on the westbound link are higher than those of the northbound links. Notice that different control policies result in different traffic flows and travel times. Therefore, it is very critical to know the control policy in effect in order to have an accurate traffic prediction.

Figure 6-7 presents a sample route choice result for the unguided users. The figure gives the probabilities of selecting path 5 for unguided users under different control policies. Path 5 starts from node 1 at Arlington Street to the westbound of Commonwealth Avenue and ends at node 28 ($1 \rightarrow 2b \rightarrow 19b \rightarrow 20a \rightarrow 21b \rightarrow 22a \rightarrow 23b \rightarrow 24a \rightarrow 25b \rightarrow 26a \rightarrow 27b \rightarrow 28$). It is the most obvious path for users traveling from node 1 to node 28 since both Arlington Street and Commonwealth Avenue are major arterial streets. As shown in Figure 6-7, the Monopoly solution assigns 100% of vehicles to this path because it is the path with minimum marginal cost. Under other solutions, less than 50% of users choosing this path.

Figure 6-8 shows the guided users' probabilities of choosing path 5. Since guided users are more sensitive than unguided users to path travel cost, guided users are more likely than unguided users to choose path 5. In addition, the choice probabilities for guided users fluctuate over time while those for unguided users are relatively stable.

Figure 6-9 presents the solution convergence situation. For each model, the objective function of the total travel time is plotted at each iteration. Notice that the results depicted in Figure 6-9 are the solutions for each *outer* iteration.

Tables 6.4–6.8 report the detailed convergence results for each control under the linear queuing delay. In each table, column 2 (Total TT) is the total travel time (minutes) at each iteration. Column 3 is the maximum of the green split absolute difference between two iterations:

$$\max_{i,m,k}\{\mid g_i^{m(I)}(k) - g_i^{m(I-1)}(k) \mid\}$$

where I is the iteration number; i is the intersection; m is the phase and k is the time interval. Column 4 is the maximum of the path flow absolute difference between two iterations:

$$\max_{r,s,p,k}\{\mid h_p^{rs(I)}(k) - h_p^{rs(I-1)}(k) \mid\}$$

Column 5 is the maximum product of the phase green split and the difference between the marginal phase delay and the minimum marginal delay (Lagrange multiplier):

$$\max_{i,m,k}\{g_i^{m(I)}(k)\,(\tilde{c}_i^{m(I)}(k) - \pi_i^{(I)})(k)\}$$

among all the phases with a positive green split, $g_i^{m(I)}(k) > 0$. This column indicates whether the first order optimality conditions (4.20)-(4.22) are satisfied. All models demonstrate good convergence results in general. Comparing columns 4 and 5 in each table, we find that the path flows converge more slowly than the green splits. This finding indicates that the DTA problem is more difficult than the DTC problem to obtain a converged solution. The finding is consistent with our analysis in Section 5.3.4 that the DTA subproblem is the bottleneck of a DTCA solution algorithm.

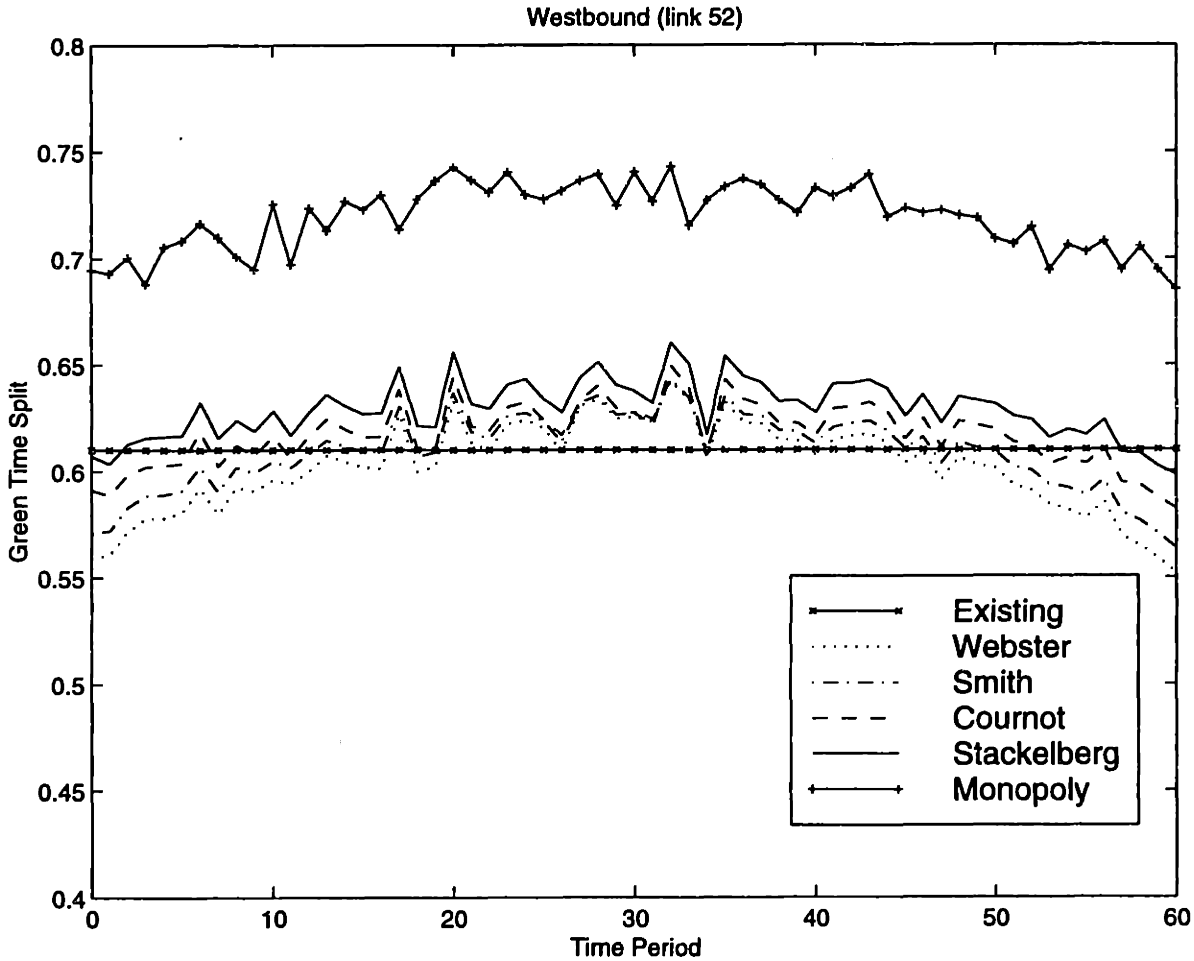

Figure 6-4: Signal settings at intersection 22: linear queuing function

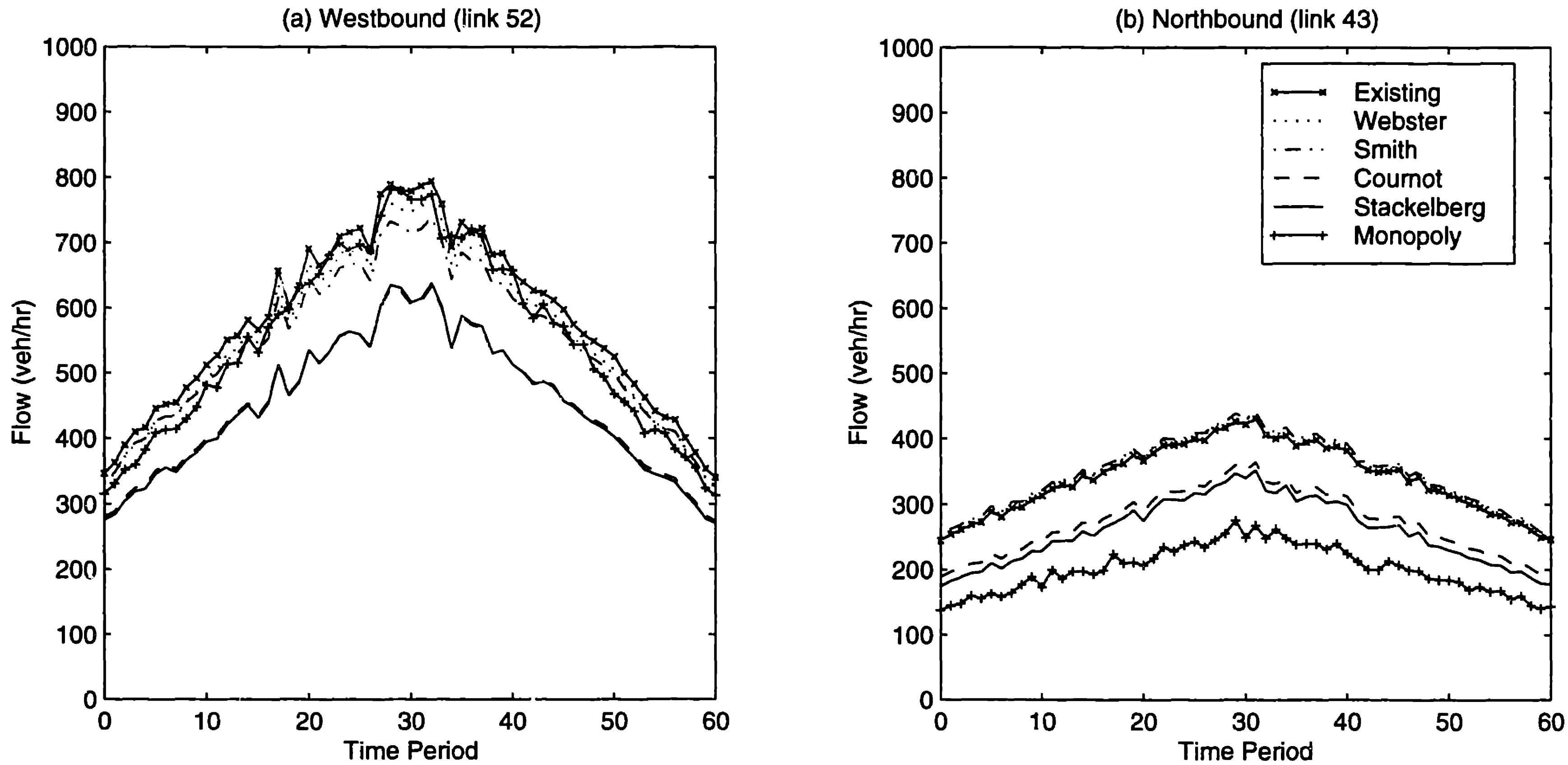

Figure 6-5: Incoming link flows at intersection 22: linear queuing function

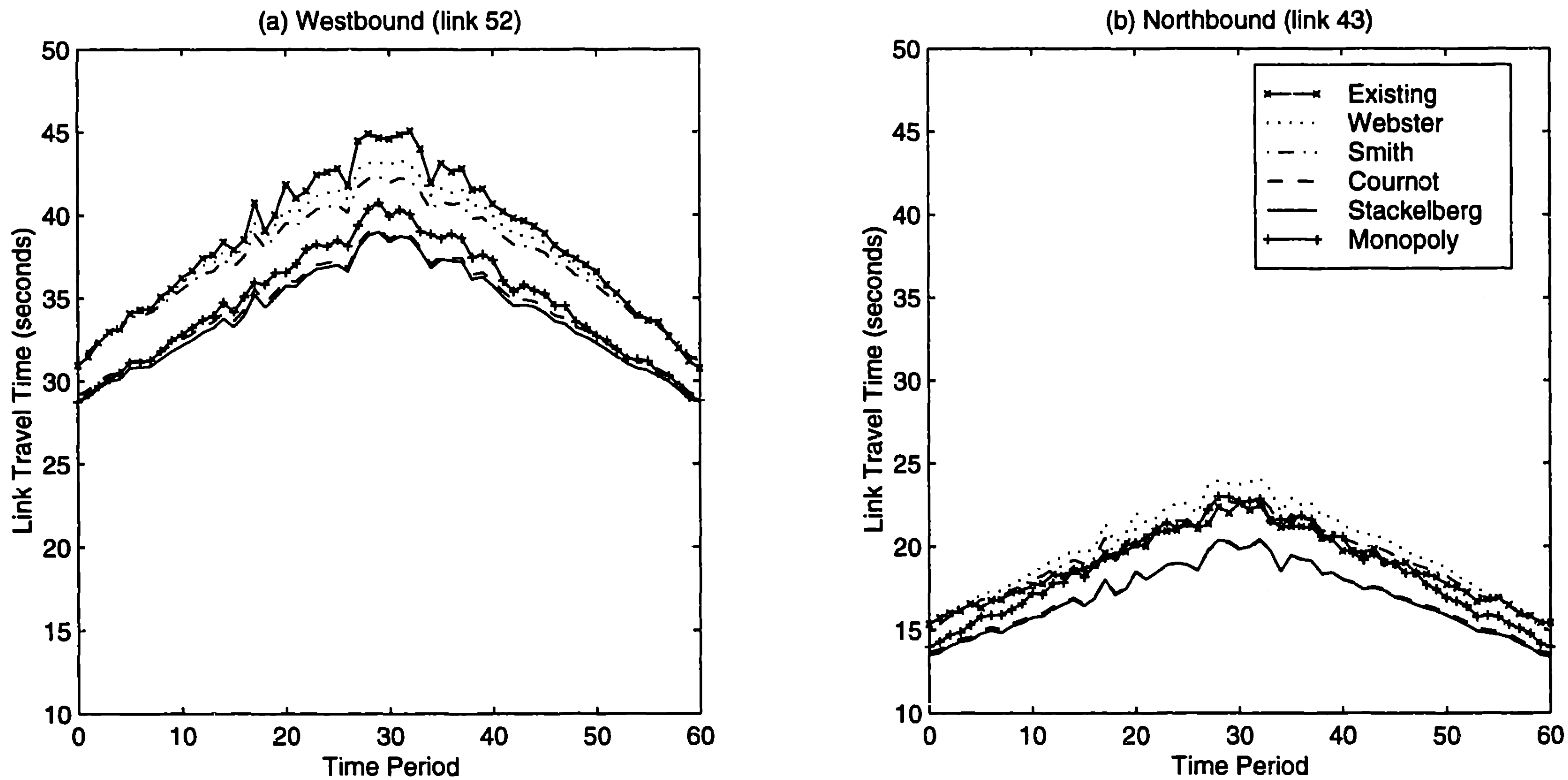

Figure 6-6: Link travel times at intersection 22: linear queuing function

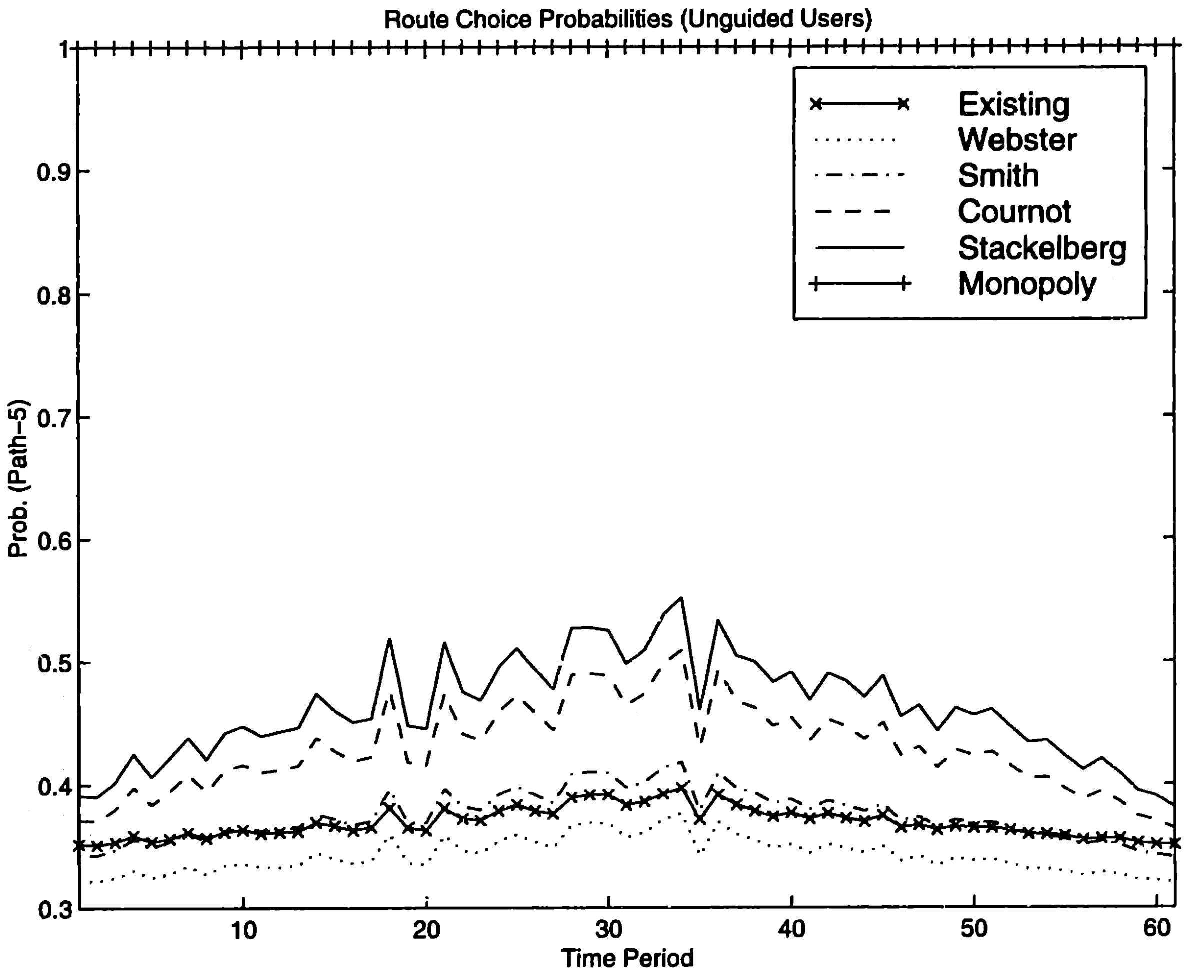

Figure 6-7: Unguided users' route choices: linear queuing function

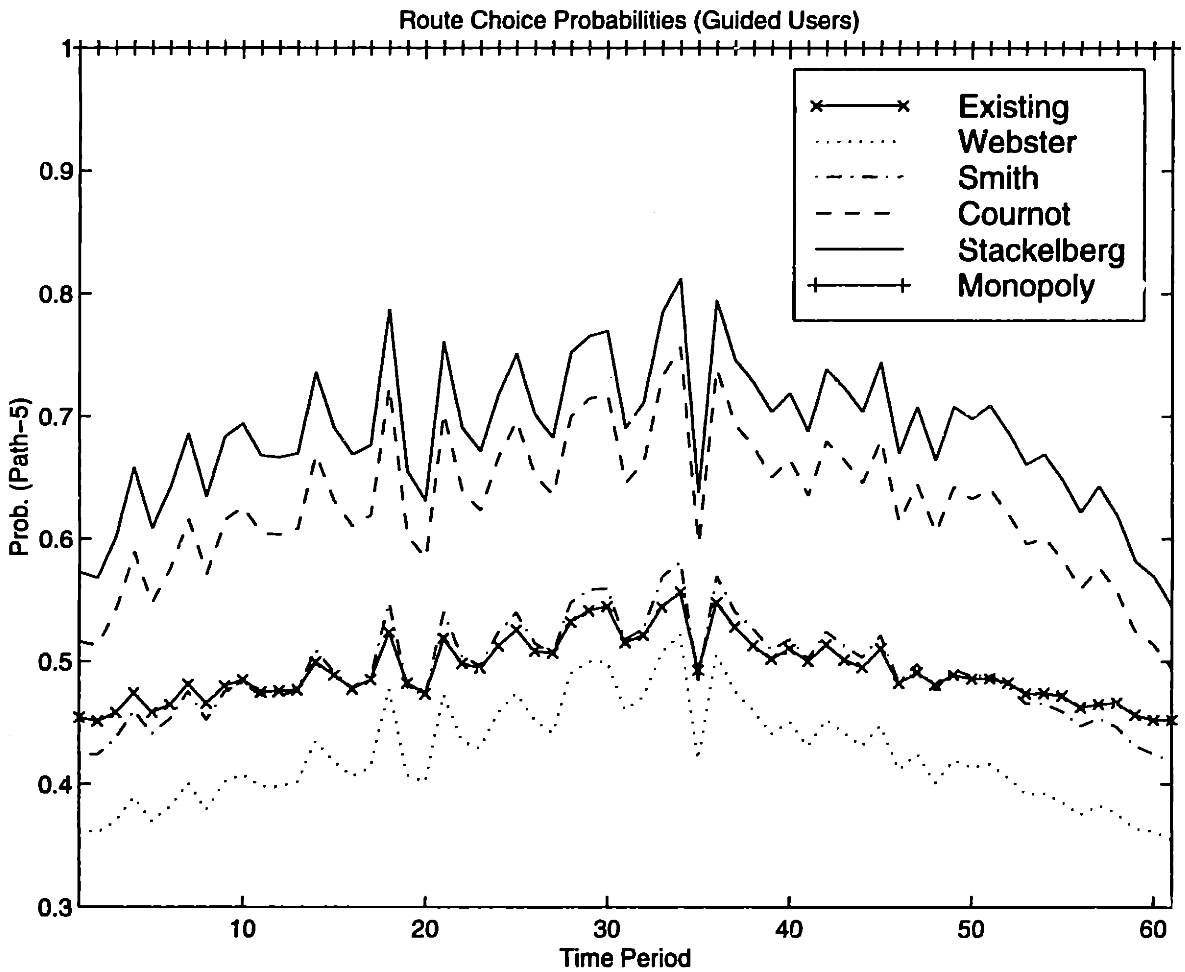

Figure 6-8: Guided users' route choices: linear queuing function

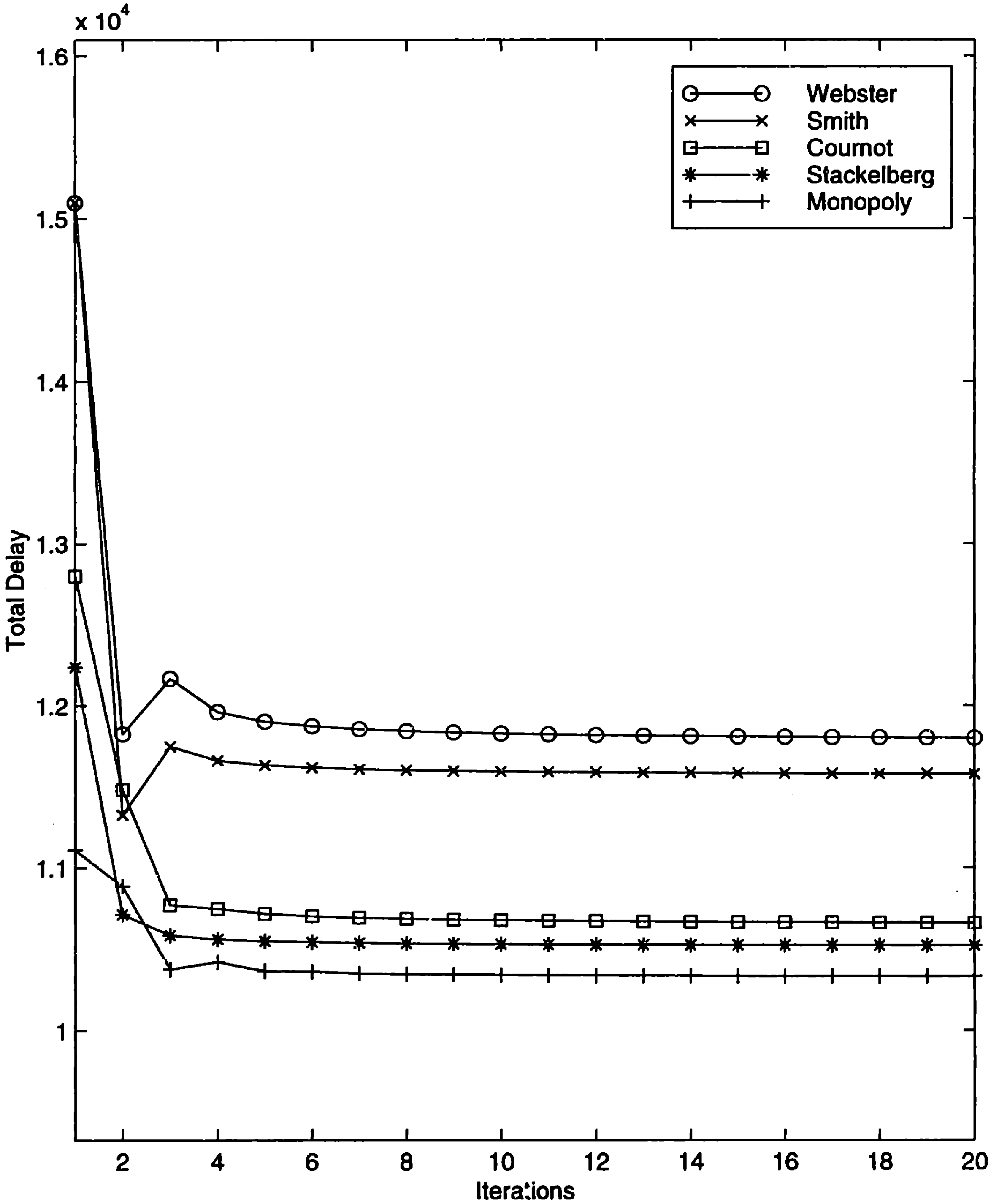

Figure 6-9: Solution convergence: linear queuing function

Table 6.4: Convergence Results for Webster Control with the Linear Queuing Function

Iteration	Total TT	$\text{Max}\{\lvert g^{(I)}-g^{(I-1)}\rvert\}$	$\text{Max}\{\lvert h^{(I)}-h^{(I-1)}\rvert\}$	$\text{Max}\{g(\tilde{c}-\pi)\}$
1	15099	0.4734	10.2783	54.8985
2	11825	0.1453	5.9489	12.1353
3	12167	0.0519	1.9737	0.9746
4	11962	0.0119	0.8567	0.1231
5	11903	0.0077	0.4401	0.0322
6	11874	0.0058	0.2638	0.0119
7	11856	0.0045	0.1756	0.0054
8	11844	0.0037	0.1256	0.0028
9	11835	0.0031	0.0945	0.0016
10	11828	0.0026	0.0738	0.0010
11	11822	0.0022	0.0594	0.0007
12	11818	0.0019	0.0489	0.0005
13	11814	0.0017	0.0409	0.0003
14	11811	0.0015	0.0348	0.0002
15	11808	0.0013	0.0300	0.0002
16	11806	0.0012	0.0262	0.0001
17	11804	0.0011	0.0230	0.0001
18	11802	0.0010	0.0204	0.0001
19	11800	0.0009	0.0182	0.0001
20	11799	0.0008	0.0164	0.0001

Table 6.5: Convergence Results for Smith's P_0 Control with the Linear Queuing Function

Iteration	Total TT	$\text{Max}\{\lvert g^{(I)} - g^{(I-1)} \rvert\}$	$\text{Max}\{\lvert h^{(I)} - h^{(I-1)} \rvert\}$	$\text{Max}\{g(\tilde{c} - \pi)\}$
1	15099	0.5000	10.2783	54.8720
2	11327	0.1469	4.1167	5.9764
3	11749	0.0383	1.1741	0.3692
4	11661	0.0122	0.4856	0.0490
5	11633	0.0080	0.2575	0.0149
6	11619	0.0056	0.1600	0.0062
7	11609	0.0042	0.1097	0.0031
8	11603	0.0033	0.0803	0.0018
9	11598	0.0026	0.0615	0.0011
10	11594	0.0022	0.0488	0.0007
11	11591	0.0019	0.0397	0.0005
12	11589	0.0017	0.0334	0.0003
13	11586	0.0016	0.0287	0.0003
14	11585	0.0015	0.0250	0.0002
15	11583	0.0013	0.0220	0.0001
16	11582	0.0012	0.0195	0.0001
17	11580	0.0011	0.0174	0.0001
18	11579	0.0011	0.0157	0.0001
19	11578	0.0010	0.0142	0.0001
20	11578	0.0009	0.0130	0.0001

Table 6.6: Convergence Results for Cournot Control with the Linear Queuing Function

Iteration	Total TT	$\text{Max}\{\lvert g^{(I)} - g^{(I-1)} \rvert\}$	$\text{Max}\{\lvert h^{(I)} - h^{(I-1)} \rvert\}$	$\text{Max}\{g(\tilde{c} - \pi)\}$
1	12800	0.4817	8.9488	12.4982
2	10749	0.1097	2.6451	0.1959
3	10770	0.0304	0.6312	0.0301
4	10729	0.0133	0.2471	0.0091
5	10710	0.0085	0.1502	0.0037
6	10698	0.0060	0.1035	0.0018
7	10690	0.0044	0.0767	0.0010
8	10685	0.0034	0.0596	0.0006
9	10680	0.0028	0.0480	0.0004
10	10677	0.0023	0.0396	0.0003
11	10674	0.0019	0.0334	0.0002
12	10671	0.0017	0.0286	0.0001
13	10669	0.0014	0.0248	0.0001
14	10667	0.0013	0.0218	0.0001
15	10666	0.0011	0.0193	0.0001
16	10665	0.0010	0.0173	0.0000
17	10663	0.0009	0.0156	0.0000
18	10662	0.0008	0.0141	0.0000
19	10661	0.0007	0.0129	0.0000
20	10660	0.0007	0.0118	0.0000

Table 6.7: Convergence Results for Stackelberg Control with the Linear Queuing Function

Iteration	Total TT	$\text{Max}\{\lvert g^{(I)} - g^{(I-1)} \rvert\}$	$\text{Max}\{\lvert h^{(I)} - h^{(I-1)} \rvert\}$	$\text{Max}\{g(\bar{c} - \pi)\}$
1	12237	0.3000	5.2562	11.9295
2	10596	0.0242	0.8367	7.7125
3	10511	0.0087	0.3250	1.0331
4	10508	0.0052	0.1774	1.3119
5	10506	0.0035	0.1128	0.2368
6	10506	0.0025	0.0784	0.1281
7	10505	0.0019	0.0579	0.0769
8	10505	0.0015	0.0449	0.0433
9	10505	0.0013	0.0360	0.0334
10	10505	0.0010	0.0296	0.0451
11	10504	0.0009	0.0249	0.0176
12	10504	0.0008	0.0212	0.0144
13	10504	0.0007	0.0183	0.0109
14	10504	0.0006	0.0160	0.0087
15	10504	0.0005	0.0141	0.0080
16	10504	0.0005	0.0126	0.0061
17	10504	0.0004	0.0113	0.0050
18	10504	0.0004	0.0102	0.0032
19	10504	0.0004	0.0093	0.0007
20	10504	0.0003	0.0084	0.0003

Table 6.8: Convergence Results for Monopoly Control with the Linear Queuing Function

Iteration	Total TT	$\mathrm{Max}\{\|g^{(I)} - g^{(I-1)}\|\}$	$\mathrm{Max}\{\|h^{(I)} - h^{(I-1)}\|\}$	$\mathrm{Max}\{g(\tilde{c} - \pi)\}$
1	11109	0.5000	2.7807	1.1929
2	10886	0.3833	3.3016	0.7713
3	10377	0.0892	1.1005	0.1033
4	10420	0.0376	1.6508	0.0312
5	10366	0.0303	0.6603	0.0237
6	10362	0.0220	0.4402	0.0128
7	10351	0.0192	0.3144	0.0077
8	10347	0.0171	0.2358	0.0043
9	10345	0.0155	0.1834	0.0033
10	10342	0.0141	0.2347	0.0015
11	10340	0.0049	0.1201	0.0008
12	10338	0.0051	0.1000	0.0004
13	10337	0.0052	0.0847	0.0001
14	10336	0.0051	0.0726	0.0000
15	10335	0.0051	0.0629	0.0000
16	10334	0.0051	0.0550	0.0000
17	10334	0.0051	0.0486	0.0000
18	10333	0.0051	0.0432	0.0000
19	10333	0.0051	0.0425	0.0000
20	10332	0.0051	0.0348	0.0000

6.2.2 Results with the HCM Delay Function

Similar results are obtained when we use the Highway Capacity Manual (HCM) delay formula as the link performance function.

Table 6.9 lists the total travel time results of the six control policies under the HCM delay function. Both Webster and Smith P_0 control policies slightly reduce the total travel time from the existing pretimed control. By contrast, both the Cournot and Stackelberg significantly reduce the total travel time from the existing solution. The reductions are 11.1% and 12.4% respectively. The gap between the Cournot and Stackelberg solutions in terms of the total travel time is 8491 - 8377 = 114 minutes or about 1.4%. Again, the Monopoly solution has the lowest total travel time.

Compared to the results in Table 6.1, the total travel times for all control policies under the HCM delay function are less than those under the linear queuing delay function. However, the optimality gaps for all non-monopoly controls are larger than those under the queuing delay function. For example, the optimality gap for the Cournot control is 4.12% in Table 6.9 while it is 3.07% in Table 6.1. Therefore, the link performance function plays a very important role in the combined dynamic traffic control-assignment (DTCA) problem.

Table 6.9: Total Travel Time Results with the HCM Function

Controls	Total Travel Time (mins)	Optimality Gap (%)
Existing	9563	17.26
Webster	9395	15.20
Smith P_0	9537	16.94
Cournot	8491	4.12
Stackelberg	8377	2.72
Monopoly	8155	0.00

Table 6.10: Link Travel Times and Speeds with the HCM Function

Controls	Avg Link TT (secs)	Change (%)	Avg Link Speed (mph)	Change (%)
Existing	40.26	0.00	7.81	0.00
Webster	39.34	-2.27	7.77	-0.42
Smith P_0	40.74	1.18	7.84	0.043
Cournot	37.57	-6.70	8.26	5.77
Stackelberg	37.18	-7.67	8.33	6.78
Monopoly	36.22	-9.77	8.57	9.80

Table 6.10 reports the average link travel time and the average link speed results. Similar to those results under the queuing delay function, the Webster control slightly improves both the mean link travel time and the speed, while the Smith P_0 control results in worse average link travel time and link speed. The three combined solutions demonstrate large reductions in the average link travel time and large increases in the average link speed.

Table 6.11 gives the average path travel time results under the HCM delay function. Again, the Webster control slightly improves the average path travel time while the Smith P_0 control results in a worse path travel time. The three combined solutions have large reductions in path travel times as well.

The signal setting at intersection 22 under the HCM delay function is given in Figure 6-10. The corresponding link flows and link travel times are shown in Figure 6-11 and Figure 6-12 respectively. The probabilities of choosing path 5 for unguided and guided users are shown in Figure 6-13 and Figure 6-14 respectively. Finally, the solution convergence is presented in Figure 6-15. These results are analogous to those under the queuing delay function.

Tables 6.12–6.16 report the detailed convergence results for each control under

Table 6.11: Path Travel Times with the HCM Delay Function

Controls	Avg Path TT (mins)	Change (%)
Existing	6.23	0.00
Webster	6.10	-2.11
Smith P_0	6.30	1.02
Cournot	5.84	-6.33
Stackelberg	5.78	-7.19
Monopoly	5.65	-9.33

the HCM delay function. Similiar to the results under the linear queuing delay, all control models show satisfactory convergence results.

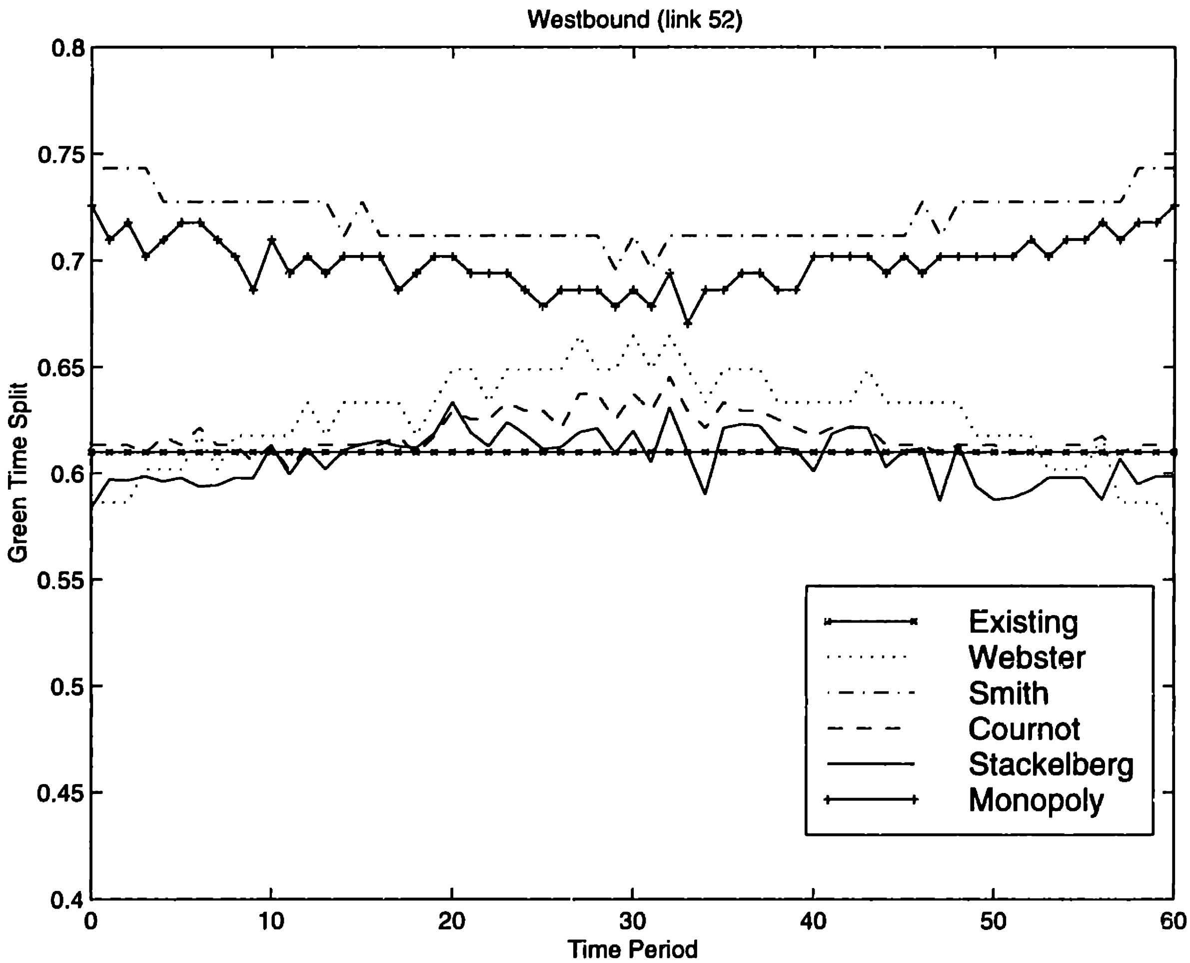

Figure 6-10: Signal settings at intersection 22: HCM function

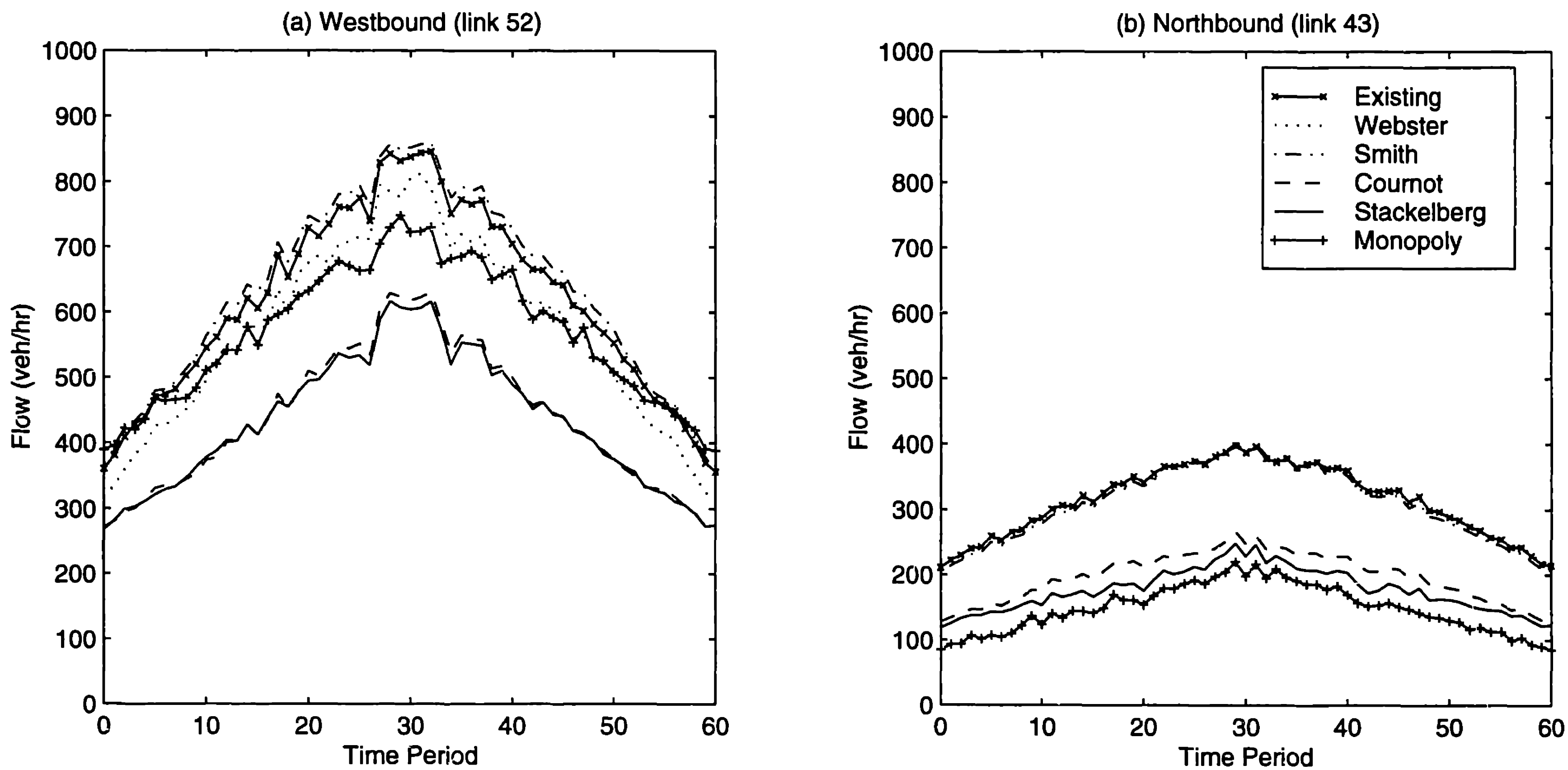

Figure 6-11: Incoming link flows at intersection 22: HCM function

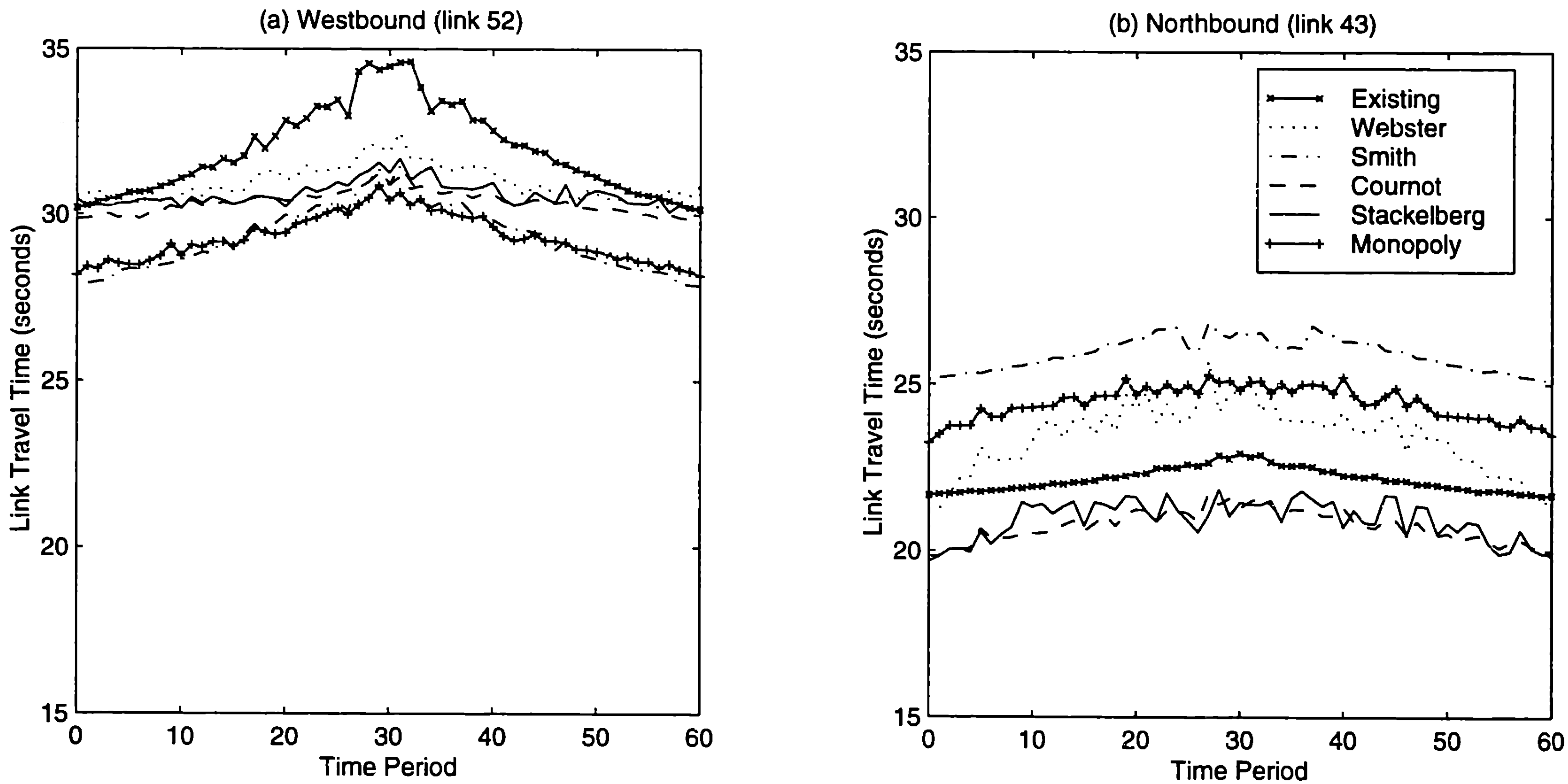

Figure 6-12: Link travel times at intersection 22: HCM function

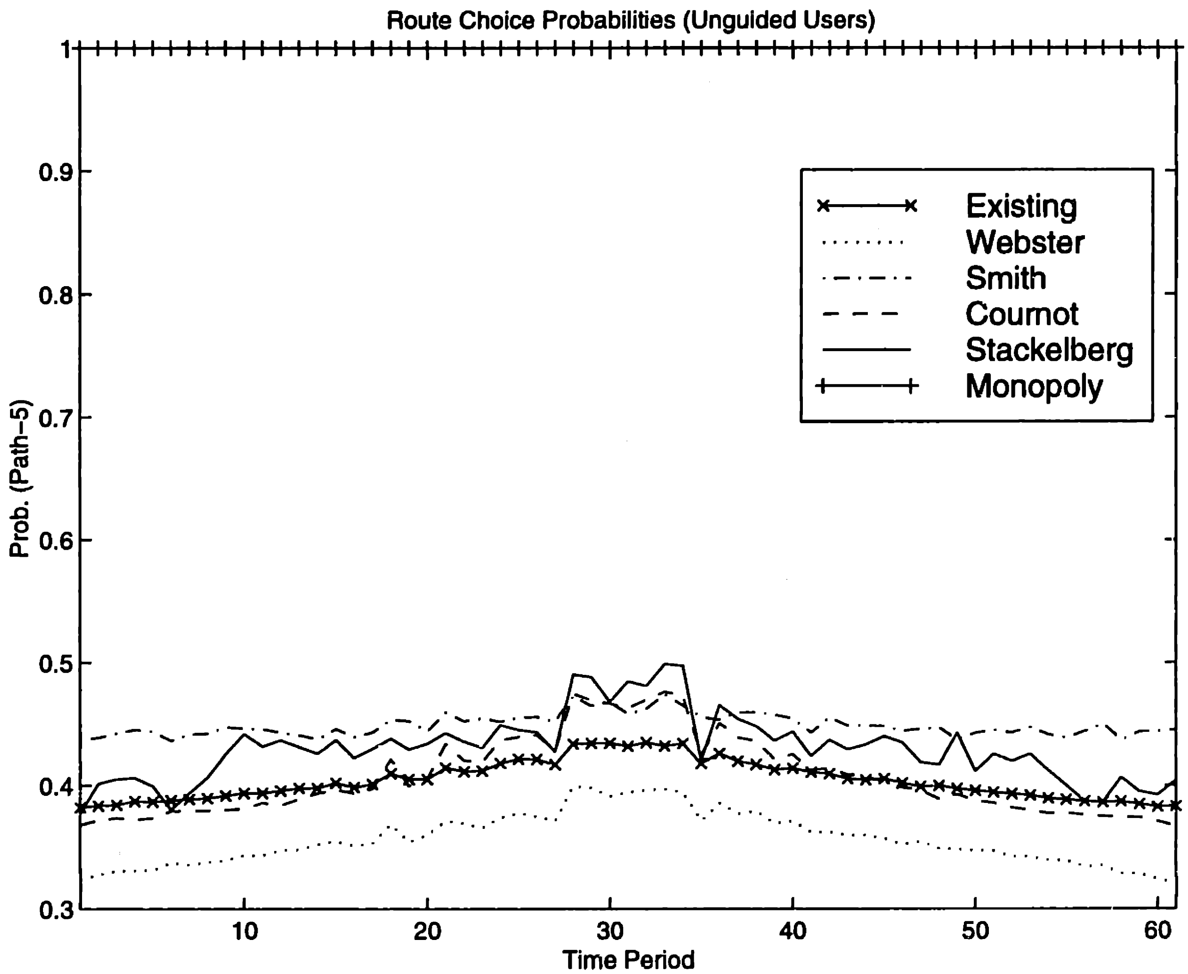

Figure 6-13: Unguided users' route choices: HCM function

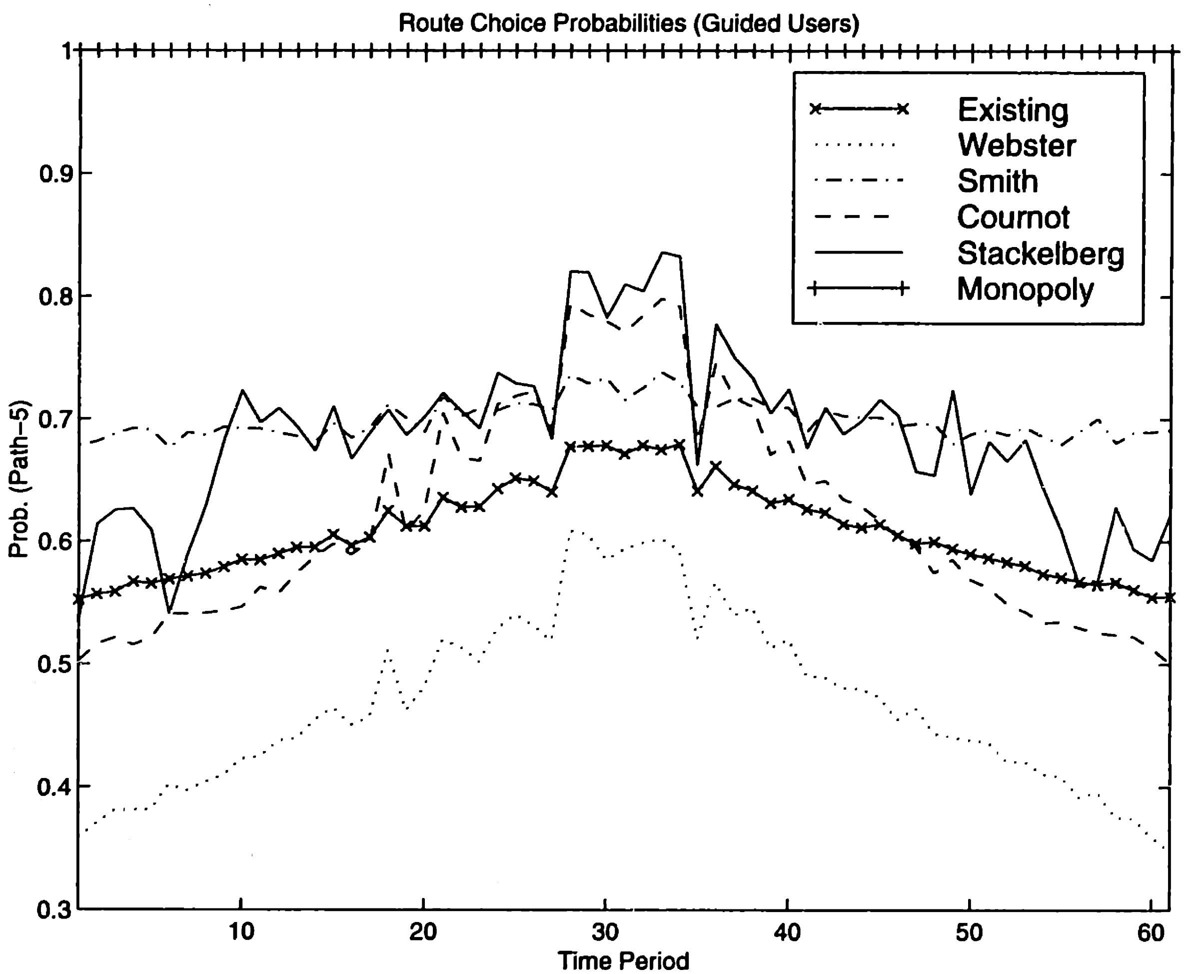

Figure 6-14: Guided users' route choices: HCM function

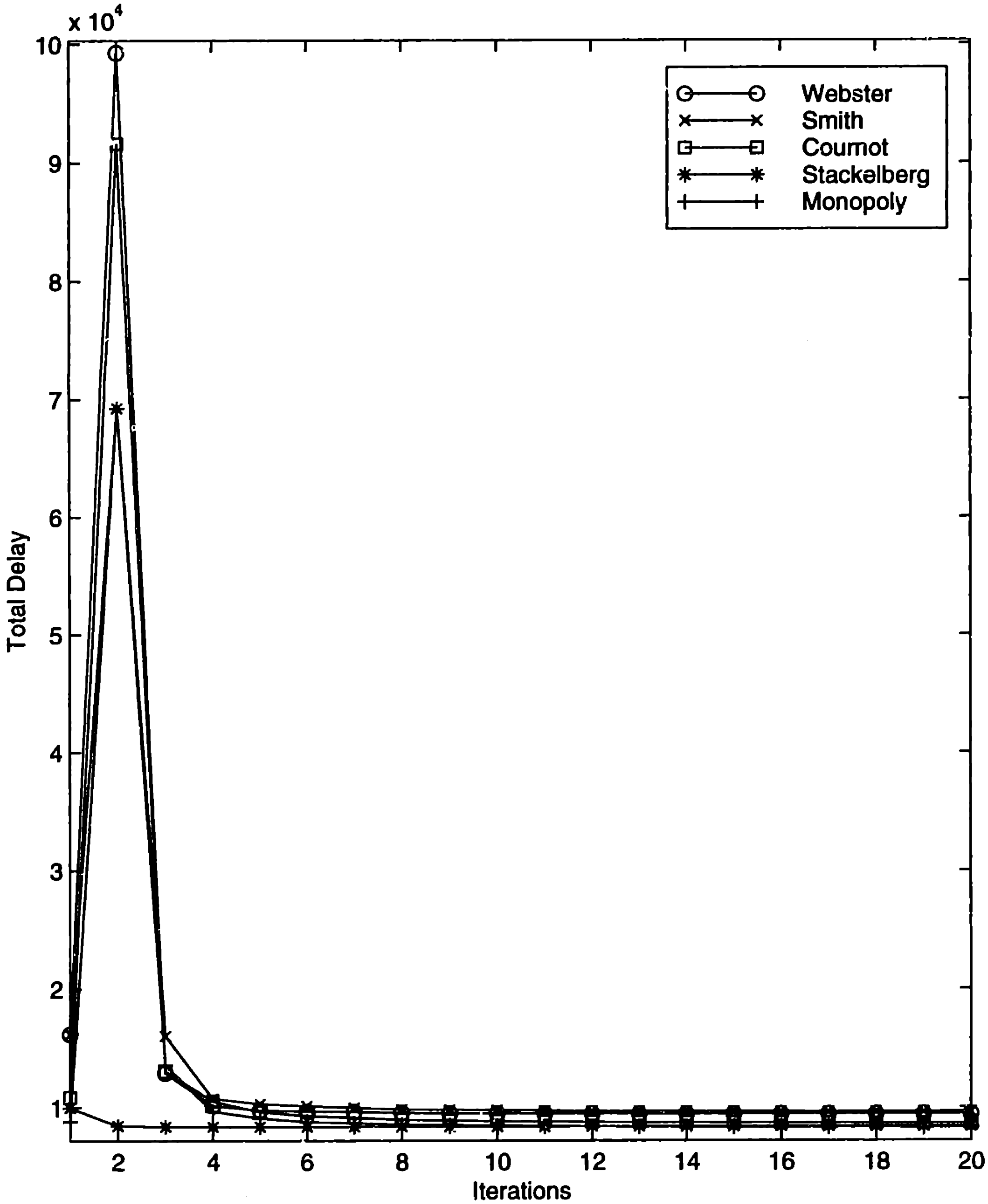

Figure 6-15: Solution convergence: HCM function

Table 6.12: Convergence Results for Webster Control with the HCM Function

Iteration	Total TT	$\text{Max}\{\lvert g^{(I)} - g^{(I-1)} \rvert\}$	$\text{Max}\{\lvert h^{(I)} - h^{(I-1)} \rvert\}$	$\text{Max}\{g(\tilde{c} - \pi)\}$
1	16162	0.5000	6.2301	3.6020
2	99297	0.4000	16.5562	4.8172
3	12909	0.2667	5.5451	0.6799
4	10131	0.1333	2.6721	0.1214
5	9729	0.1200	1.3213	0.0404
6	9651	0.1067	1.0762	0.0237
7	9607	0.0952	0.6797	0.0131
8	9536	0.0857	0.4401	0.0072
9	9501	0.0778	0.4199	0.0055
10	9484	0.0711	0.3085	0.0038
11	9468	0.0655	0.2766	0.0030
12	9456	0.0606	0.2151	0.0023
13	9447	0.0564	0.1982	0.0019
14	9442	0.0527	0.1536	0.0015
15	9438	0.0495	0.1222	0.0012
16	9433	0.0467	0.1206	0.0011
17	9424	0.0441	0.0999	0.0009
18	9428	0.0418	0.0988	0.0008
19	9422	0.0398	0.0816	0.0007
20	9424	0.0379	0.0819	0.0006

Table 6.13: Convergence Results for Smith's P_0 Control with the HCM Function

Iteration	Total TT	$\text{Max}\{\lvert g^{(I)}-g^{(I-1)}\rvert\}$	$\text{Max}\{\lvert h^{(I)}-h^{(I-1)}\rvert\}$	$\text{Max}\{g(\tilde{c}-\pi)\}$
1	16162	0.5000	6.2301	3.5275
2	69242	0.2000	8.2746	12.7052
3	16025	0.0889	1.8484	1.6071
4	10731	0.0333	0.6674	0.3106
5	10274	0.0240	0.2774	0.1025
6	10106	0.0178	0.1788	0.0537
7	9947	0.0136	0.0972	0.0250
8	9823	0.0107	0.0697	0.0142
9	9786	0.0086	0.0443	0.0099
10	9754	0.0071	0.0341	0.0067
11	9721	0.0060	0.0238	0.0046
12	9669	0.0051	0.0197	0.0033
13	9660	0.0043	0.0139	0.0025
14	9659	0.0038	0.0121	0.0020
15	9646	0.0033	0.0090	0.0015
16	9640	0.0029	0.0080	0.0014
17	9609	0.0026	0.0060	0.0010
18	9622	0.0023	0.0055	0.0009
19	9618	0.0021	0.0043	0.0008
20	9596	0.0019	0.0041	0.0007

Table 6.14: Convergence Results for Cournot Control with the HCM Function

Iteration	Total TT	$\text{Max}\{\lvert g^{(I)} - g^{(I-1)} \rvert\}$	$\text{Max}\{\lvert h^{(I)} - h^{(I-1)} \rvert\}$	$\text{Max}\{g(\tilde{c} - \pi)\}$
1	10847	0.5000	4.8987	2.1041
2	91584	0.2000	5.8097	5.5434
3	13057	0.0889	1.9973	1.1828
4	10516	0.0500	0.4823	0.1953
5	9601	0.0320	0.2328	0.0731
6	9283	0.0222	0.1278	0.0344
7	9133	0.0136	0.0790	0.0181
8	8966	0.0107	0.0514	0.0109
9	8898	0.0086	0.0357	0.0075
10	8836	0.0071	0.0255	0.0054
11	8798	0.0066	0.0191	0.0037
12	8762	0.0051	0.0145	0.0029
13	8731	0.0047	0.0114	0.0023
14	8713	0.0038	0.0090	0.0019
15	8690	0.0033	0.0073	0.0016
16	8681	0.0029	0.0060	0.0012
17	8660	0.002[illegible]	0.0050	0.0011
18	8649	0.0023	0.0042	0.0009
19	8640	0.0021	0.0035	0.0008
20	8630	0.0019	0.0030	0.0007

Table 6.15: Convergence Results for Stackelberg Control with the HCM Function

Iteration	Total TT	$\text{Max}\{\lvert g^{(I)} - g^{(I-1)} \rvert\}$	$\text{Max}\{\lvert h^{(I)} - h^{(I-1)} \rvert\}$	$\text{Max}\{g(\tilde{c} - \pi)\}$
1	9959	0.3080	1.9753	4.4157
2	8456	0.0540	0.3467	0.1704
3	8383	0.0240	0.0866	0.0288
4	8377	0.0177	0.0345	0.0095
5	8380	0.0141	0.0170	0.0043
6	8381	0.0114	0.0098	0.0023
7	8380	0.0095	0.0063	0.0014
8	8383	0.0087	0.0046	0.0009
9	8383	0.0075	0.0034	0.0007
10	8386	0.0064	0.0027	0.0005
11	8380	0.0062	0.0021	0.0004
12	8385	0.0052	0.0018	0.0003
13	8382	0.0051	0.0015	0.0002
14	8384	0.0045	0.0012	0.0002
15	8384	0.0042	0.0011	0.0002
16	8387	0.0041	0.0009	0.0001
17	8384	0.0038	0.0008	0.0001
18	8383	0.0036	0.0007	0.0001
19	8384	0.0031	0.0006	0.0001
20	8388	0.0030	0.0005	0.0001

Table 6.16: Convergence Results for Monopoly Control with the HCM Function

Iteration	Total TT	$\mathrm{Max}\{\lvert g^{(I)} - g^{(I-1)} \rvert\}$	$\mathrm{Max}\{\lvert h^{(I)} - h^{(I-1)} \rvert\}$	$\mathrm{Max}\{g(\tilde{c} - \pi)\}$
1	8835	0.5000	2.6406	1.0559
2	69219	0.2000	1.6508	1.8886
3	13555	0.0889	0.3668	0.2561
4	9654	0.0333	0.2751	0.1251
5	9093	0.0240	0.1321	0.0462
6	8777	0.0178	0.0734	0.0236
7	8671	0.0136	0.0898	0.0283
8	8561	0.0107	0.0442	0.0117
9	8511	0.0086	0.0306	0.0080
10	8471	0.0071	0.0220	0.0053
11	8429	0.0060	0.0164	0.0038
12	8394	0.0051	0.0125	0.0029
13	8363	0.0043	0.0098	0.0028
14	8345	0.0038	0.0078	0.0018
15	8323	0.0033	0.0063	0.0015
16	8317	0.0029	0.0052	0.0013
17	8304	0.0026	0.0043	0.0010
18	8300	0.0023	0.0036	0.0009
19	8285	0.0021	0.0030	0.0010
20	8276	0.0019	0.0026	0.0006

6.3 Summary

The Boston Back Bay network case study demonstrates very encouraging results. Major findings are summarized as follows:

First, the control strategies based on the three combined control-assignment models formulated in this thesis—Cournot, Stackelberg and Monopoly—show large reductions in total travel time compared to the existing pretimed control. The reductions range from 11% to 17%, depending on the link performance function used.

Second, the developed Cournot, Stackelberg and Monopoly control strategies also perform well against two other state-of-the-art control policies: the responsive Webster control and the responsive Smith P_0 control. For example, under the linear queuing delay function assumption, the Stackelberg control has a 12% improvement in the total travel time over the responsive Webster control and a 10% improvement over the responsive Smith P_0 control. Larger improvements are obtained when the HCM delay function is assumed.

Third, the assumption on the link performance function impacts the magnitude of the results. However, in the two link performance functions tested, the link performance function has minimal impacts on the performance ranking order of the control strategies.

Finally, there is a small gap between the Cournot solution and the Stackelberg solution. A Stackelberg equilibrium solution is theoretically better than a Cournot equilibrium because users' reactions to the future control settings have been taken into account when the traffic signals are optimized. This case study confirms the existence of the gap between the two equilibria. Although the best performance is represented by the Monopoly strategy, it depends on the assumption that users follow system-optimal route choices which may not be in the best interest of individual users. Therefore, the bi-level Stackelberg control strategy is the best control policy for practical application.

Chapter 7

Conclusion

This chapter summarizes the contribution of this research to the state-of-the-art of traffic control and assignment models, and provides suggestions for further research.

7.1 Contribution

This research represents an advancement of the state-of-the-art in integrating dynamic traffic control and assignment in an urban network. Specifically,

- An integrated framework is developed to combine traffic control and traffic assignment in a dynamic case. The framework allows for both a flow-based analytical modeling and a vehicle-based simulation approach. Furthermore, the framework can be implemented for both an off-line planning application and a real-time on-line operation.

- The combined dynamic traffic control-assignment (DTCA) problems are formulated for the first time as a game between a traffic authority and highway users. Three sets of models are formulated based on Cournot, Stackelberg and Monopoly games. A control strategy based on the Stackelberg model is superior to that based on a Cournot model because the traffic authority anticipates users' reactions to future control settings in solving the dynamic traffic control problem. The Monopoly strategy will have the best system performance if the users follow system-optimal route choices which in practice may not be valid. Nevertheless, a Monopoly solution represents the optimal system performance and thus serves as a benchmark for other solutions, such as Cournot and Stackelberg equilibria. Control strategies developed in existing studies, including responsive Webster control, Smith P_0 control, one-level and bi-level delay minimization controls based on combined static control-assignment models, are not optimal in the combined dynamic control-assignment problem. The bi-level Stackelberg control strategy developed in this thesis is the global optimal strategy when users choose paths to minimize their own perceived travel costs.

- Three gradient based algorithms (DTCA-1, DTCA-2 and DTCA-3) are developed to solve the DTCA problems. Algorithm DTCA-1 is an exact DTCA algorithm. At each main iteration, the algorithm solves a complete DTA subproblem and calculates the gradient of the DTC problem in exact. Algorithm DTCA-2 is a heuristic method derived from the exact algorithm of DTCA-1. Algorithm DTCA-2 consists of two inner-outer loops. In the inner loop, the two subproblems of DTC and DTA are solved in parallel. At the convergence of the inner loop, the solutions of DTC and DTA are then combined in the outer loop. Algorithm DTCA-3 is also a heuristic approximation of DTCA-1. Algorithm DTCA-3 solves a DTCA problem in one loop. In each iteration, algorithm DTCA-3 solves the two subproblems of DTC and DTA in parallel, combines their solutions and checks their convergence immediately at the end of each iteration.

- The developed control strategies based on the DTCA models are tested in a real urban network along with other control strategies in the literature. The test results show that there clearly exists a potential benefit for integrating control and assignment. A combined control-assignment solution improves the system performance not only by providing the optimal signal settings but also by giving the most accurate and consistent traffic prediction to travelers.

7.2 Further Research

The framework, models and algorithms developed in this thesis motivate a new avenue of research in the field of integrating dynamic traffic control and assignment. Possible areas for future research include the following:

Sensitivity Analysis on Users' Behavior

It should be noted that the numerical results obtained from the case study depend on the user behavior assumptions. On one extreme, if 100% of users were habitual fixed-route users, a combined control-assignment solution would not offer any benefit. On the other extreme, if 100% of users all followed deterministic dynamic user-optimal routes, both the Cournot and Stackelberg solutions would have better results than those reported in the case study. A sensitivity analysis on the classification of users and the parameters used in the C-logit route choice model should be investigated.

Efficient Solution Algorithms

The Stackelberg model is a bi-level programming problem, which is known as a *NP-Hard* problem (Jeroslow 1985). The solution algorithm developed for such a bi-level programming problem is a gradient-based method implemented in *MATLAB* code. The running time of the algorithm is much slower than real-time in a reasonably sized network. Although some efficiency can be obtained by using a *C* or *C++* implementation, the gain may be limited. Other bi-level programming solution algorithms have been developed recently, such as the penalty function approach and the interior point method. See Chen (1993) or Luo et al. (1996) for a recent review on this subject. It will be interesting to see the application of those bi-level programming algorithms to the bilevel DTCA problem. Another approach to improving solution efficiency involves an exploration of high-performance computing methods (Chabini and Ganugapati 1998), such as parallel and/or multi-thread computer implementations.

Simulation Tests

The case study results in this thesis are obtained directly from the developed models' solutions. Some simulation tests on the same network or other urban networks are needed to further evaluate the control and guidance strategies based on the combined models. For example, some scenario tests regarding the O-D demand, link performance function, guidance information, and users' route choice behavior should be conducted to understand the stability of the developed models.

Coordinated Traffic Control

In this thesis, only the green time parameters are optimized in the traffic control problem. Signal coordination must be incorporated into the signal setting process after the optimization of the green time splits. Gartner and Al-Malik (1997) present two iterative procedures to combine split and offset optimization with the traffic assignment in a static case. Both procedures use the TRANSYT program to calculate the offsets. Such iterative procedures could be extended to the dynamic case.

Congestion Pricing

Like the traffic control problem, the congestion pricing problem (Chen and Bernstein 1995) also interacts with the traffic assignment problem. A combined dynamic congestion pricing-traffic assignment problem can be formulated as a bilevel Stackelberg model in which the upper level is the dynamic optimal toll problem and the lower level is the dynamic user-optimal traffic assignment problem. Many of the models and solution algorithms developed in this thesis will still be valid for the dynamic congestion pricing problem.

Ramp Metering and Route Diversion

The framework and models in this thesis are developed for urban traffic control and management. However, they can also be extended to freeway traffic control and management. One particular application is the integration of freeway ramp metering and route diversion strategies (Chen 1996; Chen et al. 1997).

Global Optimality

In general, the Stackelberg model formulation of the DTCA problem does not have a unique optimal solution because the feasible region is non-convex. A solution that satisfies the optimality conditions is only a local optimal point. Methods must be developed to obtain a global optimal solution.

Day-To-Day Dynamics

Finally, day-to-day traffic dynamics need to be considered in a combined control-assignment environment. In fact, the iterative control-assignment process can be interpreted as the day-to-day interaction between the traffic authority and the users. However, it is implicitly assumed that users will always find their *stochastic dynamic user-optimal* route choices after the signal setting changes. The mechanism that users employ to adjust their travel decisions to the traffic control change needs to be modeled explicitly.

Bibliography

Akcelik, R. (1981). Traffic signals: capacity and timing analysis. *Australian Road Research Board Research Report*, 123.

Akcelik, R. (1988). The highway capacity manual formula for signalized intersections. *ITE Journal 58*(3), 23–27.

Akcelik, R. and M. J. Maher (1977). Route control of traffic in urban road network. *Transportation Science 11*(1), 15–24.

Al-Malik, M. and N. H. Gartner (1995). Development of a combined traffic signal control–traffic assignment model. In *Urban Traffic Networks (N. H. Gartner and G. Improta, Editors)*, pp. 155–186. Springer Verlag.

Al-Malik, M. S. (1991). *An investigation and development of a combined traffic signal control-traffic assignment model.* Ph. D. thesis, Georgia Institute of Technology.

Allsop, R. E. (1974). Some possibilities for using traffic control to influence trip distribution and route choice. In *Proceedings of the 6th International Symposium on Traffic and Transportation Theory*, Amsterdam, pp. 345–374. Elsevier.

Allsop, R. E. and J. A. Charlesworth (1977). Traffic in a signal-controlled road network: an example of different signal timings inducing different routings. *Traffic Engineering and Control 18*(5), 262–264.

Ashok, K. (1996). *Estimation and Prediction of Time-Dependent Origin-Destination Flows.* Ph. D. thesis, Massachusetts Institute of Technology.

Beckmann, M. J., C. B. McGuire, and C. B. Winsten (1956). *Studies in the Economics of Transportation.* New Haven, CT: Yale University Press.

Ben-Akiva, M. E. (1985). Dynamic network equilibrium research. *Transportation Research 19A*, 429–431.

Ben-Akiva, M. E., M. Bierlaire, J. Bottom, H. N. Koutsopoulos, and R. Mishalani (1997). Development of a route guidance generation system for real-time application. In *Proceedings of the 8th IFAC/IFIP/IFORS symposium on transportation systems*, Chania, Greece.

Ben-Akiva, M. E., E. Cascetta, and H. Gunn (1993). An on-line dynamic traffic prediction model for an inter-urban motorway network. In *Proceedings of the Second International Capri Seminar on Urban Traffic Networks*, Capri, Italy.

Ben-Akiva, M. E., E. Cascetta, H. Gunn, S. Smulders, and J. Whittaker (1994a). DYNA: A real-time monitoring and prediction system for inter-urban motorways. In *Proceedings of the First World Congress on Applications of Transport Telematics and Intelligent Vehicle-Highway Systems*, Paris, France. December, 1994.

Ben-Akiva, M. E., M. Cyna, and A. de Palma (1984). Dynamic model of peak period congestion. *Transportation Research 18B*(4), 339–355.

Ben-Akiva, M. E. and A. de Palma (1986). Some circumstances in which vehicles will reach their destinations earlier by starting later: revisited. *Transportation Science 20*(2), 52–55.

Ben-Akiva, M. E., A. de Palma, and P. Kanaroglou (1986). Dynamic model of peak period traffic congestion with elastic arrival rates. *Transportation Science 20*(2), 164–181.

Ben-Akiva, M. E., A. de Palma, and I. Kaysi (1991). Dynamic network models and driver information systems. *Transportation Research 25A*(5), 251–266.

Ben-Akiva, M. E., H. Koutsopoulos, and A. Mukundan (1994b). A dynamic traffic model system for ATMS/ATIS operations. *IVHS Journal 2*, 1–19.

Ben-Akiva, M. E. and S. R. Lerman (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand.* Cambridge, MA: MIT Press.

Bertsekas, D. P. (1995). *Nonlinear Porgamming.* Belmont, MA: Athena Scientific.

Boyce, D. E., B. Ran, and L. J. LeBlanc (1995). Solving an instantaneous dynamic user-optimal route choice model. *Transportation Science 29*, 128–142.

Cantarella, G. E., G. Improta, and A. Sforza (1991). Iterative procedure for equilibrium network traffic signal setting. *Transportation Research 24A*, 241–249.

Carey, M. (1991). Optimal time-varying flows on congested networks. *Operations Research 35*(1), 58–69.

Carey, M. (1992). Nonconvexity of the dynamic traffic assignment problem. *Transportation Research 26B*, 127–133.

Cascetta, E. and G. Cantarella (1991). A day-to-day and within-day dynamic stochastic assignment model. *Transportation Research 25A*, 277–291.

Cascetta, E. and G. Cantarella (1995). Dynamic processes and equilibrium in transportation networks: towards a unifying theory. *Transportation Science 29*, 305–329.

Cascetta, E., K. Di Gangi, K. Lindveld, and P. Mijjer (1994). A dynamic traffic assignment model for real time applications. In *Proceedings of the First World Congress on Applications of Transport Telematics and Intelligent Vehicle-Highway Systems*, Paris, France. December, 1994.

Cascetta, E., A. Nuzzolo, F. Russo, and A. Vitetta (1996). A modified logit route choice model overcoming path overlapping problems: specification and some calibration results for interurban networks. In *Proceedings of the 13th International Symposium on Transportation and Traffic Theory*, Lyon, France, July 24-26, 1996, pp. 697–711.

Cascetta, E., F. Russo, and A. Vitetta (1997). Stochastic user equilibrium assignment with explicit path enumeration: comparison of models and algorithms. In *Proceedings of the 8th International Federation of Automatic Control (IFAC) Symposium on Transportation Systems*, Chania, Greece, June 16-18, 1997, pp. 1078–1084.

Chabini, I. (1997a). Discrete dynamic shortest path problems in transportation applications: complexity and algorithms with optimal run time. To appear in *Transportation Research Record*.

Chabini, I. (1997b). A new short path algorithm for discrete dynamic networks. In *Proceedings of the 8th International Federation of Automatic Control (IFAC) Symposium on Transportation Systems*, Chania, Greece, June 16-18, 1997, pp. 551–556.

Chabini, I. and S. Ganugapati (1998). High performance computing of dynamic network models. Presented in the CORS/INFORMS Meeting, Montreal, Canada, April 26-29, 1998.

Chabini, I. and Y. He (1997). An analytical approach to dynamic network loading: formulation, algorithms, and computer implementations. Submitted to *Transportation Science*.

Chabini, I. and Y. He (1998a). A flow-based approach to dynamic traffic assignment. Presented in the CORS/INFORMS Meeting, Montreal, Canada, April 26-29, 1998.

Chabini, I. and Y. He (1998b). A flow-based approach to the dynamic traffic assignment problem: formulations, algorithms, and computer implementations. To be presented in the Third Triennial Symposium on Transportation Analysis (TRISTAN III), San Juan, Puerto Rico, June 17-23, 1998.

Chen, H.-K. and C.-F. Hsueh (1997). Combining signal timing plan and dynamic traffic assignment. Presented at the *76th Annual Transportation Research Board Meeting*. Washington, D.C., January 12-16, 1997.

Chen, O. J. (1996). *A Dynamic Traffic Control Model For Real-Time Freeway Operations.* MST Thesis, Massachusetts Institute of Technology.

Chen, O. J. and M. E. Ben-Akiva (1998a). Dynamic traffic control and assignment: a game-theoretic approach. Presented in the *77th Annual Meeting of the Transportation Research Board.* Washington,D.C., January 11-15, 1998.

Chen, O. J. and M. E. Ben-Akiva (1998b). Dynamic traffic control and traffic assignment. Presented in the CORS/INFORMS Meeting, Montreal, Canada, April 26-29, 1998.

Chen, O. J. and M. E. Ben-Akiva (1998c). Game-theoretic formulations of the interaction between dynamic traffic control and dynamic traffic control assignment. To appear in *Transportation Research Record.*

Chen, O. J. and D. Bernstein (1995). AM/PM congestion pricing with a single toll plaza. *Transportation Research Record 1498*, 23–31.

Chen, O. J., A. F. Hotz, and M. E. Ben-Akiva (1997). Development and evaluation of a dynamic traffic control model for real-time freeway operations. In *Proceedings of the 8th International Federation of Automatic Control (IFAC) Symposium on Transportation Systems*, Chania, Greece, June 16-18, 1997, pp. 1162–1168.

Chen, Y. (1993). *Bilevel Programming Problems: Analysis, Algorithms and Applications.* Ph. D. thesis, Université de Montréal.

Chiou, S.-W. (1997). Optimization of area traffic control subject to user equilibrium traffic assignment. In *Proceedings of the 25th European Transport Forum, Seminar F*, Brunel University, September, 1997, pp. 53–64.

Cournot, A. (1838). *Recherches sur les Principes Mathématiques de la Théorie des Richesses.* English edition: *Researches into Mathematical Principles of the Theory of Wealth.* Macmillan, New York. 1897.

Dafermos, S. C. (1971). An extended traffic assignment model with application to two-way traffic. *Transportation Science 5*(4), 366–389.

Dafermos, S. C. (1980). Traffic equilibrium and variational inequalities. *Transportation Science 14*, 42–54.

Daganzo, C. (1995). Properties of link travel time functions under dynamic loads. *Transportation Research 29B*, 95–98.

Daganzo, C. and Y. Sheffi (1977). On stochastic models of traffic assignment. *Transportation Science 11*(3), 253–274.

Di Taranto, C. and V. Mauro (1989). UTOPIA. In *Proceedings of 6th IFAC/IFIP/IFORS Symposium on Transportation*, Paris, France.

Dickson, T. J. (1981). A note on traffic assignment and signal timings in a signal-controlled road network. *Transportation Science 15B*, 267–271.

Drissi-Kaitouni, O. and A. Hameda-Benchekroun (1992). A dynamic traffic assignment model and solution algorithm. *Transportation Science 26*, 119–128.

FHWA (1985). *Traffic Control Systems Handbook.* U.S. Department of Transportation, Federal Highway Administration, Washington, D.C.

Fisk, C. S. (1984). Game theory and transportation systems modeling. *Transportation Research 18B*, 301–313.

Frank, M. and P. Wolfe (1956). An algorithm for quadratic programming. *Naval research logistics quarterly 3*(1-2), 95–110.

Friesz, T. L. (1985). Transportation network equilibrium, design and aggregation: key development and research opportunities. *Transportation Research 19A*, 413–427.

Friesz, T. L., D. Bernstein, T. E. Smith, and B. Wie (1993). A variational inequality formulation of the dynamic network user equilibrium problem. *Operations Research 41*, 179–191.

Friesz, T. L., F. J. Luque, R. L. Tobin, and W. B. (1989). Dynamic network traffic assignment considered as a continuous time optimal control problem. *Operations Research 37*, 893–901.

Fudenberg, D. (1991). *Game Theory.* Cambridge, MA: MIT Press.

Gartner, N. H. (1976). Area traffic control and network equilibrium. In *Proceedings of the International Symposium on Traffic Equilibrium Methods (edited by Florian, M.A.)*, pp. 274–297. Springer Verlag.

Gartner, N. H. (1977). Influencing traffic equilibrium in a network by control measures. *Transportation Research Record 644*, 125–129.

Gartner, N. H. (1983). OPAC: A demand-responsive strategy for traffic signal control. *Transportation Research Record 906.*

Gartner, N. H. and M. Al-Malik (1996). Combined model for signal control and route choice in urban traffic networks. *Transportation Research Record 1554*, 27–35.

Gartner, N. H. and M. Al-Malik (1997). Combined control and route assignment in traffic signal networks. In *Proceedings of the 8th International Federation of Automatic Control (IFAC) Symposium on Transportation Systems*, Chania, Greece, June 16-18, 1997, pp. 634–639.

Gartner, N. H., S. Gershwin, J. D. C. Little, and P. Ross (1980). Pilot study of computer-based urban traffic management. *Transportation Research 14B*(6), 453–462.

Gartner, N. H., J. D. C. Little, and H. Gabbay (1975). Optimization of traffic signal settings by mixed-integer linear programming–Part I: the network coordination problem, and Part II: the network synchronization problem. *Transportation Science 9*, 321–363.

Gartner, N. H. and C. Stamatiadis (1996). Framework for the integration of dynamic traffic assignment with real-time control. In *Proceedings of the 3rd Annual World Congress on Intelligent Transportation Systems, Orlando, Florida.*

Gartner, N. H., C. Stamatiadis, and P. Tarnoff (1995). Development of advanced traffic signal control strategies for ITS: a multi-level design. *Transportation Research Record 1494*, 98–105.

Gartner, N. H., P. Tarnoff, and C. Andrews (1991). Evaluation of optimized policies for adaptive control strategy. *Transportation Research Record 1324*, 105–114.

Gershwin, S., J. D. C. Little, and N. H. Gartner (1978). Computer-assisted traffic engineering using assignment, optimal signal setting, and modal split. Report No. DOT-TSC-RSPA-78-10, U.S. Department of Transportation, Research and Special Programs Administration, Washington, DC.

Gibbons, R. (1992). *Game Theory For Applied Economists.* Princeton, NJ: Princeton University Press.

HCM (1994). *Highway Capacity Manual.* Special Report 209, Third Edition, Transportation Research Board, Washington, D.C.

He, Y. (1997). *A Flow-Based Approach To The Dynamic Traffic Assignment Problem: Formulations, Algorithms, And Computer Implementations.* MST Thesis, Massachusetts Institute of Technology.

Henry, J. J. and J. L. Farges (1989). PRODYN. In *Proceedings of 6th IFAC/IFIP/IFORS Symposium on Transportation*, Paris, France.

Heydecker, B. G. (1983). Some consequences of detailed junction modeling in road traffic assignment. *Transportation Science 17*, 263–281.

Heydecker, B. G. and T. K. Khoo (1990). The equilibrium network design problem. In *Proceedings of AIRO'90 Conference on Models and Methods for Decision Support*, pp. 587–602.

Hunt, P. B., D. I. Robertson, R. D. Bretherton, and R. Winton (1981). SCOOT-a traffic responsive method of coordinating signals. In *TRRL Report LR 1014*. Transport and Road Research Laboratory, Crowthorne, U.K.

Janson, B. (1991). Dynamic traffic assignment for urban networks. *Transportation Research 25B*, 143–161.

Janson, B. and J. Robles (1991). Dynamic traffic assignment with arrival time costs. In *Proceedings of 12th International Symposium on Transportation and Traffic Theory*, pp. 127–146. Elsevier Science, Amsterdam, The Netherlands.

Jauffred, F. and D. Bernstein (1996). An alternative formulation of the simultaneous route and departure-time choice equilibrium problem. *Transportation Research 4C*(6), 339–357.

Jayakrishnan, R., H. Mahmassani, and T.-Y. Hu (1994). An evaluation tool for advanced traffic information and management systems in urban networks. *Transportation Research 2C*, 129–147.

Jeroslow, R. (1985). The polynomial hierarchy and a simple model for competitive analysis. *Mathematical Programming 32*, 146–164.

Kreps, D. (1990). *Game Theory and Economic Modeling.* Oxford, UK: Oxford University Press.

Leonard, D., P. Grower, and N. Taylor (1989). CONTRAM: structure of the model. *TRRL Research Report, No. 178.*

Lieberman, E. B. and J. L. Woo (1982). SIGOP-III user's manual. *U.S. Department of Transportation, Federal Highway Administration Report No. FHWA-IP-82-A.*

Little, J. D. C., M. Kelson, and N. H. Gartner (1981). MAXBAND: a program for setting signals on arteries and triangular networks. *Transportation Research Record 795*, 40–46.

Lowrie, P. R. (1981). SCATS: the Sydney Co-ordinated Adaptive Traffic System. In *International Conference on Road Traffic Signaling.* Crowthorne, U.K.

Luo, Z.-Q., P. S. Pang, and D. Ralph (1996). *Mathematical Programs with Equilibrium Constraints.* Cambridge, UK: Cambridge University Press.

MacGowan, J. and I. J. Fullerton (1979-1980). Development and testing of advanced control strategies in the urban traffic control system. *Public Roads 43*(2, 3, 4).

Maher, M. J. and R. Akcelik (1975). The re-distributional effects of area traffic control policy. *Traffic Engineering and Control 16*(9), 383–385.

Maher, M. J. and R. Akcelik (1977). Route control-simulation experiments. *Transportation Science 11*(1), 25–31.

Mahmassani, H., T.-Y. Hu, and R. Jayakrishnan (1992). Dynamic traffic assignment and simulation for advanced network informatics (DYNASMART). In *Proceedings of the Second International Capri Seminar on Urban Traffic Networks*. Capri, Italy.

Marcotte, P. (1986). Network design problem with congestion effects: A case of bilevel programming. *Mathematical Programming 34*, 142–162.

Merchant, D. K. and G. L. Nemhauser (1978a). A model and an algorithm for the dynamic traffic assignment problems. *Transportation Science 12*, 183–199.

Merchant, D. K. and G. L. Nemhauser (1978b). Optimality conditions for a dynamic traffic assignment model. *Transportation Science 12*, 200–207.

Nagurney, A. (1993). *Network economics: a variational inequality approach*. Dordrecht: Kluwer Academic Publishers.

Nash, J. F. (1950). Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences, U.S.A. 36*, 48–49.

Nash, J. F. (1951). Noncooperative games. *Annals of Mathematics 54*, 286–295.

Ran, B. and D. E. Boyce (1996). *Modeling dynamic transportation networks: an intelligent transportation system oriented approach*. Berlin: Springer.

Ran, B., R. W. Hall, and D. E. Boyce (1996). A link-based variational inequality model for dynamic departure time/route choice. *Transportation Research 30B*, 31–46.

Ran, B., I. Y. Li, and W. B. D. Soetopo (1997). An analytical path-based multi-class dynamic traffic assignment model. In *Preprints of the 76th Annual Meeting of the Transportation Research Board*, Washington D. C.

Ran, B. and T. Shimazaki (1989a). Dynamic user equilibrium traffic assignment for congested transportation networks. *Presented at the Fifth World Conference on Transport Research.*

Ran, B. and T. Shimazaki (1989b). A general model and algorithm for dynamic traffic assignment problems. In *Proceedings of Fifth World Conference on Transport Research*, pp. 463–477. Yokohama, Japan.

Robertson, D. I. (1969). TRANSYT method for area traffic control. *Traffic Engineering and Control 11*(10), 259–275.

Sheffi, Y. (1984). *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods.* Englewood Cliffs, NJ.: Prentice-Hall, Inc.

Sheffi, Y. and W. B. Powell (1983). Optimal signal settings over transportation networks. *Journal of Transportation Engineering 109*(6), 824–839.

Sim, A. G. (1979). The Sydney Co-ordinated Adaptive Traffic System. In *Proceedings of Engineering Foundation Conference on Research Priorities in Computer Control of Urban Traffic Systems*, New York, pp. 12–27. ASCE.

Smith, M. J. (1979). Traffic control and route choice: a simple example. *Transportation Research 13B*(4), 289–294.

Smith, M. J. (1981a). The existence of an equilibrium solution to the traffic assignment problem when there are junction interactions. *Transportation Research 15B*(6), 443–452.

Smith, M. J. (1981b). Properties of a traffic control policy which ensure the existence of traffic equilibrium consistent with the policy. *Transportation Research 15B*(6), 453–462.

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.

Smith, M. J. and M. O. Ghali (1990). The dynamics of traffic assignment and traffic control: a theoretical study. *Transportation Research 24B*(6), 409–422.

Smith, M. J. and T. Van Vuren (1993). Traffic equilibrium with responsive traffic control. *Transportation Science 27*(2), 118–132.

Tan, H.-N., S. Gershwin, and M. Athans (1979). Hybrid optimization in urban traffic networks. In *Report No. DOT-TSC-RSPA-79-7.* U.S. Department of Transportation, Research and Special Programs Administration, Washington, DC.

von Stackelberg, H. (1934). *Marktform und Gleichgewicht.* Vienna: Julius Springer. English edition: *The Theory of the Market Economy.* Oxford University Press, Oxford, England, 1952.

Vythoulkas, P. (1990). A dynamic stochastic assignment model for the analysis of general networks. *Transportation Research 24B*(6), 253–272.

Wardrop, J. G. (1952). Some theoretical aspects of road traffic research. *Proceedings of the Institute of Civil Engineers, Part II 1*, 325–378.

Webster, F. (1958). Traffic signal settings. *Road Research Technical Paper No. 39.*

Wie, B.-W. (1989). Dynamic system optimal traffic assignment on congested multidestination networks. *Paper Presented at the 5th World Conference on Transport Research, Yokahoma, Japan.*

Wie, B.-W., T. L. Friesz, and R. L. Tobin (1990). Dynamic user optimal traffic assignment on congested multidestination networks. *Transportation Research 24B*, 431–442.

Wie, B.-W., R. L. Tobin, T. L. Friesz, and D. Bernstein (1995). A discrete time, nested cost operator approach to the dynamic network user equilibrium problem. *Transportation Science 29*, 79–92.

Wu, J. H., Y. Chen, and M. Florian (1995). The continuous dynamic network loading problem: A mathematical formulation and solution method. Technical report, CRT-95-25, Centre de Recherche sur les Transports, Université de Montréal.

Xu, Y., J. H. Wu, M. Florian, P. Marcotte, , and D. L. Zhu (1996). New advances in the continuous dynamic network loading problem. Technical report, CRT-96-26, Centre de Recherche sur les Transports, Université de Montréal.

Yang, H. and S. Yagar (1994). Traffic assignment and traffic control in general freeway arterial corridor systems. *Transportation Research 28B*(6), 463–486.

Yang, H. and S. Yagar (1995). Traffic assignment and traffic control in saturated road networks. *Transportation Research 29A*(2), 125–139.

Yang, Q. (1997). *A Simulation Laboratory for Evaluation of Dynamic Traffic Management Systems.* Ph. D. thesis, Massachusetts Institute of Technology.

Yang, Q. and H. N. Koutsopoulos (1996). A microscopic traffic simulator for evaluation of dynamic traffic management systems. *Transportation Research 4*(3), 113–129.

Appendix A

Summary of Notations

The notations used throughout the thesis are defined below:

Sets:

$\mathcal{A}$	$=$	set of links;
$\mathcal{N}$	$=$	set of nodes (intersections);
$\mathcal{R} \subseteq \mathcal{N}$	$=$	set of origin nodes;
$\mathcal{S} \subseteq \mathcal{N}$	$=$	set of destination nodes;
$\mathcal{P}^{rs}$	$=$	set of all paths from node r to node s;
$\mathcal{A}(i)$	$=$	set of links whose tail node is i (i's outgoing links);
$\mathcal{B}(i)$	$=$	set of links whose head node is i (i's incoming links);
$\mathcal{M}_i$	$=$	set of phases for intersection i;
$\mathcal{G}$	$=$	set of all feasible green time splits;
$\mathcal{H}$	$=$	set of all feasible path flows;

Index Variables:

t = time;

$i \in \mathcal{N}$ = node (intersection) index;

$a \in \mathcal{A}$ = link index;

$p \in \mathcal{P}$ = path index;

$r \in \mathcal{R}$ = origin index;

$s \in \mathcal{S}$ = destination index;

$m \in \mathcal{M}_i$ = signal phasing index for each intersection i;

Traffic Flow Variables:

$x_{ap}^{rs}(t)$ = number of vehicles on link a from r to s along path p at time t;

$u_{ap}^{rs}(t)$ = entry flow on link a from r to s along path p at time t;

$v_{ap}^{rs}(t)$ = exit flow on link a from r to s along path p at time t;

$U_{ap}^{rs}(t)$ = cumulative number of vehicles entering link a from r to s along path p by time t;

$V_{ap}^{rs}(t)$ = cumulative number of vehicles leaving link a from r to s along path p by time t;

$P_{np}^{rs}(t)$ = probability of choosing path p of O-D pair (r, s) for class-n users departing from origin r to destination s at time t;

$h_{np}^{rs}(t)$ = flow of class n users departed from r to s along path p at time t;

$h_{p}^{rs}(t)$ = $\sum_n h_{np}^{rs}(t)$, total flow departed from r to s along path p at time t;

$D_{n}^{rs}(t)$ = demand of class n users from origin r to destination s at time t;

$D^{rs}(t)$ = $\sum_n D_{n}^{rs}(t)$, total demand from origin r to destination s at time t;

Signal Control Variables:

$C_i(t)$ = cycle length for intersection i at time t;

$G_i^m(t)$ = green time for phase m of intersection i at time t;

$G_i^{m,min}$ = minimum green time for phase m of intersection i;

$g_i^m(t)$ = green split for phase m of intersection i at time t;

$l_i^m(t)$ = loss time for phase m of intersection i at time t;

$$\theta_{ia}^{m}(t) = \begin{cases} 1 & \text{if link } a \text{ is in phase } m \text{ at time } t \\ 0 & \text{otherwise} \end{cases}$$

Cost Functions:

$\tau_a(t)$ = link a travel cost for flows entering the link at time t;

$s_a(t)$ = link exit time for a vehicle entering link a at time t;

$c_p^{rs}(t)$ = actual travel cost for path p between O-D pair (r,s) for flows departing from origin r at time t;

$\epsilon_p^{rs}(t)$ = random error of perceived travel cost for path p between O-D pair (r,s) for flows departing from origin r at time t;

$c_p^{rs}(t) + \epsilon_p^{rs}(t)$ = perceived travel cost for path p between O-D pair (r,s) for flows departing from origin r at time t;

$\tilde{\tau}_a(t)$ = $\frac{\partial[u_a(t)\tau_a(t)]}{\partial g_a(t)}$, marginal link delay with respect to green split for link a at time t;

$\tilde{c}_i^m(t)$ = $\sum_a \tilde{\tau}_a(t)\theta_{ia}^m(t)$, marginal phase delay with respect to green split for phase m of intersection i at time t;

$\pi_i(t)$ = $\min_{m\in\mathcal{M}_i} \tilde{c}_i^m(t)$, minimal marginal phase delay for intersection i at time t;

$E_{np}^{rs}(t)$ = $\left[h_{np}^{rs}(t) - D_n^{rs}(t)\, P_{np}^{rs}(t)\right] \frac{\partial c_p^{rs}(t)}{\partial h_{np}^{rs}(t)}$, auxiliary path cost function for class n users.

Appendix B

Abbreviation

ATIS Advanced Traveler Information Systems

ATMS Advanced Traffic Management Systems

DNL Dynamic Network Loading

DSO Dynamic System Optimum

DTA Dynamic Traffic Assignment

DTC Dynamic Traffic Control

DTCA Dynamic Traffic Control-Assignment

DTMS Dynamic Traffic Management Systems

DUO Dynamic User Optimum

HCM Highway Capacity Manual

FIFO First-In-First-Out

ITS Intelligent Transportation Systems

KKT Karush-Kuhn-Tucker Conditions

MP Mathematical Program

MPEC Mathematical Program with Equilibrium Constraints

NLP Non-Linear Program

O-D Origin-destination.

SDUO Stochastic Dynamic User Optimum

UTCS Urban Traffic Control System

VI Variational Inequality