A Dynamic Traffic Control Model for Real Time Freeway Operations

by

Owen Jianwen Chen

Bachelor of Engineering in Management & Systems Science
University of Science and Technology of China, Hefei, China
(1991)

Master of Arts in Economics
Vanderbilt University, Nashville, TN
(1993)

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

MASTER OF SCIENCE IN TRANSPORTATION
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
May 1996

©Massachusetts Institute of Technology 1996. All rights reserved.

Author_________________________________________________________

Department of Civil and Environmental Engineering
May 24, 1996

Certified by_____________________________________________________

Moshe E. Ben-Akiva
Professor, Department of Civil and Environmental Engineering
Thesis Supervisor

Certified by_____________________________________________________

Anthony F. Hotz
Research Associate, Center for Transportation Studies
Thesis Supervisor

Certified by_____________________________________________________

Joseph M. Sussman
Chairman, Departmental Committee on Graduate Studies
A Dynamic Traffic Control Model for Real Time Freeway Operations

by

Owen Jianwen Chen

Submitted to the Department of Civil and Environmental Engineering on May 24, 1996,
in partial fulfillment of the requirements
for the degree of Master of Science in Transportation

Abstract

Advances in information and communication technologies have made possible new developments in dynamic traffic control systems. A number of such control systems are deployed around the world under the context of the Intelligent Transportation Systems (ITS). This thesis presents a new dynamic traffic control model that can be applied in a real-time freeway traffic management environment.

The model consists of four components: state estimation, O-D prediction, local control and area-wide control. These components are integrated in a hierarchical structure and combines feedback control (local control module) with predictive control (area-wide control module).

A new dynamic optimal control algorithm is implemented with area-wide control module. This new area-wide control model is formulated and solved as a mathematical program with a linear objective function and non-linear constraints. Compared to existing algorithms, it has the following advantages: (i) it captures the true traffic flow conservation by modeling the traffic flows according to their origin-destination pairs; the differentiation of flows by O-D pairs allows the implementation of those control strategies that are specific to the O-D information, such as route guidance; (ii) it uses a general non-linear speed-density function; (iii) it optimizes the control and predicts the traffic simultaneously; and (iv) it may be implemented on-line in a rolling horizon approach.

The proposed model and associated algorithms are tested in a freeway corridor control with ramp metering in a microscopic simulation environment. The results show that the bilevel algorithm which combines the local feedback control and the new area-wide dynamic optimal control works best in all tested scenarios.

Thesis jointly supervised by:
Moshe E. Ben-Akiva Professor, Department of Civil and Environmental Engineering
Anthony F. Hotz, Research Associate, Center for Transportation Studies
To

Di Wu
Acknowledgments

I acknowledge with deep gratitude and appreciation the guidance provided by my advisors, Prof. Moshe Ben-Akiva and Mr. Tony Hotz. They provided a lot of encouragement and help for which I will be forever grateful.

Special thanks go Dr. Rabi Mishalani for his valuable time and guidance. His support and criticism pulled me through many difficulties. This thesis can not be done without the innumerable prolonged discussions that I had with Rabi and Tony. I am also grateful to Prof. David Bernstein at Princeton for his guidance and attention he provided when he was at MIT.

Many thanks go to Qi Yang for his superb simulator, without which I could not have tested the model in this thesis. I thank my other fellows at the Intelligent Transportation Research Program, specially Peter Welch, Kazi Ahmed, Hariharan Subramanian, Kalidas Ashok and Winston Guo.

I would like to thank the Massachusetts Highway Department and Bechtel/Parsons Brinckerhoff company for their financial support through the Central Artery/Tunnel Project at MIT.

Finally, I would like to thank my wife, Di Wu, for her understanding, encouragement, patience, support and love.
Contents

Abstract ................................................................................................................................................. 3

Acknowledgments ............................................................................................................................... 7

Contents .................................................................................................................................................. 9

List of Figures ........................................................................................................................................ 11

List of Tables ......................................................................................................................................... 13

Chapter 1 Introduction .......................................................................................................................... 14

1.1 The Congestion Problem ............................................................................................................... 14
1.2 Opportunities for New Traffic Control Strategies ...................................................................... 16
1.3 Research Contribution .................................................................................................................. 18
1.4 Thesis Outline ............................................................................................................................... 19

Chapter 2 Hierarchical Feedback Control Systems .............................................................................. 21

2.1 Control System Properties and Structure ...................................................................................... 21
  2.1.1 Centralized and Distributed Systems .................................................................................... 21
  2.1.2 Static and Dynamic Systems ............................................................................................... 24
  2.1.3 Open Loop and Closed Loop Systems ............................................................................... 24
  2.1.4 Reactive and Predictive Systems ......................................................................................... 26

2.2 Dynamic Traffic Control System ................................................................................................. 27
  2.2.1 Surveillance Devices ............................................................................................................ 29
  2.2.2 Control Devices .................................................................................................................. 29
  2.2.3 Dynamic Traffic Control Model ......................................................................................... 29

2.3 Proposed Hierarchical Feedback Traffic Control System ............................................................... 32
  2.3.1 Overall Structure ............................................................................................................... 32
  2.3.2 State Estimation ................................................................................................................ 33
  2.3.3 OD Prediction .................................................................................................................... 35
  2.3.4 Area-wide Control ............................................................................................................ 35
  2.3.4 Local Control .................................................................................................................... 38

2.4 Summary .......................................................................................................................................... 38

Chapter 3 Literature Review on Traffic Control Models ....................................................................... 39

3.1 Area-Wide Traffic Control ............................................................................................................. 39
  3.1.1 Static Control Model ........................................................................................................... 40
  3.1.2 Sequential Model ............................................................................................................... 42
  3.1.3 Dynamic Analytical Model ............................................................................................... 44
  3.1.4 Summary of the Analytical Control Models ...................................................................... 47
  3.1.5 Dynamic Simulation Model ............................................................................................... 47
List of Figures

Figure 1.1 Percentage of Urban Peak-Hour Travel Under Congestion .................. 15
Figure 2.1 Centralized System ........................................................................... 22
Figure 2.2 Distributed System ........................................................................... 23
Figure 2.3 Hierarchical System ......................................................................... 23
Figure 2.4 Open-Loop Control System ................................................................ 25
Figure 2.5 Closed-Loop Control System ............................................................. 25
Figure 2.6 Dynamic Traffic Control in A Closed Loop System ......................... 28
Figure 2.7 A Hierarchical Feedback System for the Dynamic Traffic Control ......... 33
Figure 2.8 A rolling horizon approach to area-wide control ............................... 35
Figure 3.1 A portion of freeway with on-ramp control ......................................... 40
Figure 3.2 Simulation-based dynamic control model ........................................... 48
Figure 3.3 A local traffic process ....................................................................... 49
Figure 4.1 An example of freeway network ......................................................... 57
Figure 4.2 Network representation for the example freeway ............................... 57
Figure 4.3 Mainline Flow Conservation ............................................................. 77
Figure 4.4 Speed-Density Relationship ............................................................... 63
Figure 4.5 On-ramp Queuing Conservation ......................................................... 64
Figure 4.6 Dynamic Traffic State Evolution ......................................................... 66
Figure 4.7 Flow Chart of the Solution Algorithm ................................................ 70
Figure 5.1 Evaluation Framework for a Dynamic Traffic Control System .......... 74
Figure 5.2 Implementation of the Dynamic Traffic Control Evaluation ............ 76
Figure 5.3 The Central Artery/Ted Williams Tunnel project network ............... 78
Figure 5.4 Case Study Network (I93 South Bound) ........................................... 79
Figure 5.5 Network Representation for I-93 ....................................................... 80
Figure 5.6 Low Demand ................................................................................... 82
Figure 5.7 Moderate Demand .......................................................................... 83
Figure 5.8 High Demand .................................................................................. 84
List of Tables

Table 3.1 Summary of Freeway Traffic Control Models ................................................. 54
Table 5.1 Simulation Results of Total Throughput ......................................................... 105
Table 5.2 O-D Travel Times --Low Demand ................................................................. 108
Table 5.3 O-D Travel Times --Moderate Demand ......................................................... 109
Table 5.4 O-D Travel Times --High Demand ............................................................... 110
Table 5.5 Simulation Results of Total Travel Time ....................................................... 112
Chapter 1
Introduction

Advances in information and communication technologies have made possible new developments in dynamic traffic control systems. A number of such control systems are deployed around the world under the context of Intelligent Transportation Systems (ITS). This thesis presents a real-time, dynamic traffic control system that uses both local and area-wide information to improve traffic performance in a freeway network.

This chapter motivates this research by first reviewing the problem of traffic congestion and discussing opportunities for new control strategies. The research contribution is then outlined and the chapter concludes by laying out the structure of the thesis.

1.1 The Congestion Problem

Travel by private automobile has grown enormously during the past three decades. Roadway expansion, on the other hand, has slowed. The result is that the highway congestion has increased substantially. Rush-hour conditions in U.S. metropolitan areas often extend throughout the day. According to Highway Statistics (Federal Highway Administration 1990), the peak hour travel under congested conditions\(^1\) on urban interstate freeways has reached 69 percent in 1990, up from 41 percent in 1975 (see Figure 1.1). The annual urban freeway delay is estimated at 2 billion vehicle hours.

\(^1\) The congested condition defined by Highway Statistics is that the flow-capacity ratio is greater than or equal to 80%.
Furthermore, the increasing congestion has also reduced highway safety. Each year the U.S. pays an extraordinary cost of 41,000 deaths and 5 million injuries on our highways. Additionally, traffic accidents cost the U.S. an estimated $70 billion in lost wages and other direct costs annually (ITS America 1995).

Unfortunately, congestion also results in lost productivity and damage to the environment. The loss in productivity due to normal traffic delay is estimated at additional $100 billion per year (ITS America 1995) and approximately 66 percent of all carbon monoxide (CO)
emissions are generated by automobiles. Energy efficiency is another concern. Fuel consumption in the U.S. has almost tripled between 1955 and 1990 and 63 percent of the fuel is consumed by automobiles (FHWA 1990).

1.2 Opportunities for New Traffic Control Strategies

The traditional approach of building new roads to solve the congestion problem is very costly, disruptive to existing traffic, opposed by numerous groups on environmental grounds, and constrained by available land. Consequently, alternative solutions aimed at effective utilization of existing systems have gained more attention. Advanced traffic management is one of these alternatives and has been identified as one of the most important user services by the National ITS Program Plan (ITS America 1995). Traffic control that aims at better managing the movement of traffic on streets and highways can alleviate congestion by:

- utilizing capacity efficiently at intersections, on-ramps and off-ramps;
- smoothing the flow of traffic on highways and thus reducing recurrent congestion and non-recurrent accidents;
- guiding travelers along the best routes to their destinations;
- detecting and clearing incidents quickly;
- shifting traffic demand from peak to off-peak;
- giving preference to high occupancy vehicles;
- reducing vehicle idle time and thus reducing emissions and fuel consumption.

Advances in surveillance, control, communications and computing technologies bring new opportunities for designing innovative control strategies to realize the above control objectives. These opportunities include:
• **Comprehensive real-time surveillance systems**
  Comprehensive traffic surveillance systems and communication systems, such as vehicle sensors, closed circuit television (CCTV), over-height vehicle detectors, weigh-in-motion devices, automatic vehicle location (AVL) and vehicle/roadside communications (VRC) devices, provide rich real-time information about the current traffic conditions.

• **Sophisticated data processing techniques**
  Real-time data processing techniques, that includes real-time data fusion, traffic state estimation and dynamic demand estimation, can provide real-time traffic state (such as speed, density, volume) and dynamic origin-destination demand information.

• **Advanced traffic control technologies**
  Advanced traffic control devices such as variable message signs, lane use signs, variable speed limit signs and even in-vehicle guidance devices provide greater flexibility in delivering traffic control and information to drivers.

• **Advances in traffic flow modeling**
  Recent advances in traffic flow modeling such as dynamic traffic assignment and dynamic traffic simulation provide better analytical tools for traffic control design.

• **Advances in modern control theory**
  Modern control theory has advanced significantly in the areas of dynamic optimal control and adaptive control. Such advances provide a theoretic foundation for a robust traffic control design.

• ** Powerful computation environment**
Today's powerful computers, including both workstations and personal computers, provide the necessary memory and speed to perform the complex computations needed for real-time traffic control and off-time optimization.

Given the availability of these advanced technologies, what is still needed is the development of new traffic control strategies that exploits these technologies to better manage traffic flow. Several such advanced control strategies are currently under development, including dynamic optimal ramp and mainline metering (Papageorgiou et al. 1990 & 1991, Zhang et al. 1995), real-time variable speed control (Smulders 1990), real-time incident detection and management (Parkany 1993), dynamic route guidance (Ben-Akiva et al. 1991, Ran and Boyce 1994) and congestion pricing (Chen and Bernstein 1995). This theses presents a new dynamic control model that uses many of the above technologies to better manage traffic in a freeway network.

1.3 Research Contribution

The primary objective of this research is to develop an advanced dynamic traffic control model for real-time freeway operations. The proposed model is a hierarchical system that manages traffic at both the local level and the area-wide level. It has the following properties:

- is functionally and/or geometrically distributed;
- accepts real-time surveillance data;
- estimates and predicts O-D flows;
- captures real-time traffic dynamics;
- estimates current network conditions;
- predicts future traffic conditions;
- optimizes control strategies based on current and future conditions;
- provides a stable control over all expected traffic conditions;
• is able to run faster than real-time for on-line implementation.

First, a control theoretic approach is used to formalize the hierarchical structure of the model. Then, a new area-wide dynamic optimal control model, is formulated and solved. Finally, the proposed model is tested in a real-time simulation environment for the dynamic freeway corridor control for the Central Artery/Ted Williams Tunnel network in Boston.

1.4 Thesis Outline

The remaining chapters of this thesis are organized as follows.

Chapter 2 presents the hierarchical structure of the proposed dynamic traffic control model. Some important control system properties are reviewed first. Then four components of the proposed model—state estimation, O-D prediction, local control and area-wide control are discussed.

Chapter 3 reviews research on freeway traffic control models developed during the past three decades. Four area-wide control models and two local control models are discussed in detail. Some experiences and limitations of those models are also identified.

Chapter 4 presents a new area-wide, dynamic optimal control model. The dynamic optimal control problem is formulated as a non-linear program. The traffic flow dynamics in the mainline and the ramps is captured in detail by modeling time-dependent origin-destination flows. A solution algorithm is also presented.

Chapter 5 applies the proposed model to the ramp metering control in the Central Artery/Ted William Tunnel network. Several control algorithms are compared and tested in a microscopic traffic simulator. These include local linear-quadratic feedback control,
area-wide dynamic optimal control, and a bilevel control that combines both local and area controls.

Chapter 6 concludes this thesis and gives directions for future research.
Chapter 2
Hierarchical Feedback Control Systems

Modern traffic control systems consist of various control devices, traffic detectors, communication links and control models that process surveillance data and generates control commands for the control devices. Traffic control can be applied either locally or area-wide. This chapter reviews the fundamental approach to traffic control design using a hierarchical feedback approach that integrates local and area-wide control.

2.1 Control System Properties and Structure

This section reviews some basic properties, definitions and structure of general control systems. Control Systems can be classified as centralized or distributed, static or dynamic, open-looped or closed-looped, linear or non-linear, discrete or continuous, and adaptive or predictive. Each of these systems is reviewed below.

2.1.1 Centralized and Distributed Systems

Control systems can be classified as centralized or distributed, according to their functional architecture. As shown in Figure 2.1, a centralized system consists of a single process that coordinates a set of controllers over the network and optimizes the control parameters in the central computer based on the overall system objective. Since all inputs
and outputs are processed in the control center, a centralized system requires intensive communication links between the control center and all field controllers. In contrast, a distributed system consists of many distributed processes as shown in Figure 2.2. Since each process is independent of the others, such are robust to failures of individual components, i.e., if one controller fails, the integrity of the system remains intact. Moreover, a distributed system processes information more efficiently and generally responds more quickly to a local system changes than does a centralize system. However, such distributed systems are independent and hence cannot be coordinated to achieve a system-wide objective.

Hierarchical systems combine both centralized and distributed structures (Papageorgiou 1983 and Hotz et al. 1992) as shown in Figure 2.3. A hierarchical control system can improve the tasks of processing and communicating large amounts of data. Information processing is subsequently hierarchically distributed. Such systems have demonstrated high reliability and robustness (Hotz et al. 1992). Moreover, a functionally distributed system favors a geographically distributed system since some dense communications links can be performed locally, hence reducing communications requirements over longer distances. Such systems are ideally suited for traffic control systems and provided the functional structure for the hierarchical control system presented later.

![Centralized System Diagram](image)

**Figure 2.1 Centralized System**
Figure 2.2 Distributed System

Figure 2.3 Hierarchical System
2.1.2 Static and Dynamic Systems

A system is static if it does not change over a sufficiently long time interval. A static system deals best with average situations over a certain period of time. There are two major disadvantages of a static traffic control system. First, the system does not respond to changes in traffic conditions. The control parameters are independent of time and hence independent of real traffic conditions. Second, the traffic flow in a static system is assumed to be in a steady-state, or equilibrium condition. This assumption is easily violated in real traffic flow scenarios.

In a dynamic traffic control system, the traffic flow is assumed to be dynamic and the control output is also time-dependent. In most cases a dynamic system is superior to a static one. However, a dynamic control system requires extra hardware and software, compared to a static system. For example, a dynamic system requires surveillance detectors to collect current traffic information, a central computer to process such information and generate the control, and communications links among the surveillance detectors, the central computer and the control devices.

2.1.3 Open Loop and Closed Loop Systems

Traffic control systems can also be categorized as open-loop or closed-loop. As shown in Figure 2.4, a system is open-loop if the control input is independent of the system output. That is, when applied to traffic control, the control commands are not dependent on current traffic conditions. Its control plans are calculated off-line according to historical data. Typical open-loop traffic control systems are fixed-time control and time-of-day control. Fixed-time control implements one single constant plan while time-of-day control can have multiple plans that are stored in the controller’s memory and selected by time-of-day criterion (such as peak period vs. off-peak period). Most of existing traffic control systems for both freeways and surface streets are still using either fixed-time or time-of-day control. As inferred from Figure 2.4, open-loop systems do not require on-line surveillance or communication links.
Figure 2.4  Open-Loop Control System

Figure 2.5  Closed-Loop Control System
In contrast, the control input to a closed loop system is a function of the system output. Figure 2.5 illustrates the closed-loop control concept. By definition, a closed loop system requires measurements of the current system state. The real-time control plans are computed on-line based on surveillance data. Consequently, closed-loop traffic control systems respond to the changes in traffic conditions in real-time.

Open loop control works best when the relationship between the input to the control loop and its output, i.e., the system model, is precisely known and no disturbances to the system are present. Certainly, neither of these conditions are generally satisfied for traffic control applications. Traffic models used for off-line control optimization are not precise and disturbances to traffic flow are almost always present. Consequently, traffic control designs using pre-timed algorithms are always suboptimal, even if they have been "optimized" off-line. One way of improving the control performance of such system is to implement closed-loop control. Closed loop controllers are more robust in a stochastic environment because they adapt to changing traffic conditions.

2.1.4 Reactive and Predictive Systems

The closed loop control described earlier is generally reactive, in the sense that an error must occur (or build-up) before any control action is taken. The control input is only adjusted backward to actual state vectors. This means that the control is only responsive to what has happened. This type of control is also known as proportional-integral control in control theory. When applying such algorithms to traffic systems, it requires that a congestion must appear before a change in traffic control is initiated. Hence, these control algorithms will usually render suboptimal performance.

The reactive control system can generally be improved by including an "anticipatory" term in the control algorithm. The idea is to adjust the control not only to what has happened but also to what is predicted to happen in the future. For example, if there is an accident at a downstream, we can predict that it will have a congestion effect on upstream traffic. If the traffic control responds earlier before the congestion reaches upstream, the impact
can be minimized at the lowest level. In general, a predictive control system first predicts future conditions and then optimizes the control based on measured and predicted conditions. Since the prediction depends current and future control inputs and the optimal control also depends on the predicted future conditions, the consistency between the prediction and the control must be satisfied. In other words, the optimal predictive control is a fixed point problem. This makes the search for an optimal and consistent predictive control very challenging.

2.2 Dynamic Traffic Control System

The simplest form of a closed loop dynamic traffic control system is shown in Figure 2.6. The system consists of surveillance devices, traffic network, control devices and a dynamic traffic control model. The state of the traffic system is governed by exogenous disturbances (such as traffic demand variations and accidents) and control devices. This system state is monitored by surveillance devices that provide inputs to the dynamic traffic control (DTC) model. The DTC model in turn generates control commands that are based on the surveillance data. This control loop is implemented on-line so that the control model can respond in real-time to current and predicted disturbances. The feedback between the traffic system and the DTC model is a key feature in this architecture. The proposed hierarchical DTC model is based on this structure and presented in the next section. However, before proceeding, the components shown in Figure 2.6 are described below in details.
Figure 2.6 Dynamic Traffic Control in A Closed Loop System
2.2.1 Surveillance Devices

Surveillance devices or sensors include point sensors, point-to-point sensors and area sensors. Point sensors collect non-spatial information such as occupancy, vehicle counts, vehicle classification, headways and point speeds. Examples of point sensors are inductive loops, fiber-optic pressure sensors, magnetometers, ultrasonic detectors, and infrared detectors. Point-to-point sensors collect data from the same vehicle at different points on the network. Such data includes origin and destination, vehicle identification, travel time, and space mean speed. Vehicle/roadside communication (VRC) systems are the most common point-to-point sensor. Finally area sensors measure area parameters such as density and queue length. An example of area sensors is a video camera.

2.2.2 Control Devices

Important control devices implemented in this closed loop structure include mainline meters, ramp meters, traffic signals, variable message signs, variable speed limit signs, variable lane use signs and toll booths. These devices can provide either control or information.

2.2.3 Dynamic Traffic Control Model

The traffic control model represents the traffic network which includes both the network geometry, i.e., freeways and arterials, and vehicles. For modeling convenience, the roadway network is usually divided into many homogenous segments. As a result, the space is discrete. Vehicles, on the other hand, are treated continuously as flows, though they can have many types such as high performance cars, low performance cars, high occupancy vehicles, heavy vehicles, etc. Since vehicles or flows change over time, the traffic network is also dynamic. Again, for simplicity, time is treated to be discrete. The time horizon is divided into many time intervals with each having equal length. However,
the length of the time interval may vary from model to model. Key to this component is a traffic flow model that is used to describe the movement of vehicles over the network.

Finally, the control model is designed to satisfy a specific or a set of specific control functions for the system. A control function is a mechanism by which a specific design goal may be achieved. The control function is implemented by choosing a control device and its control algorithm that governs the behavior of the control function. The algorithm in turn determines the type of input data (surveillance data) that is needed and generates the control commands that set the states of the control device. Basic control functions in for freeway management include

- Route diversion
- Access control
- Mainline control
- Arterial control

Each of these functions is briefly described below.

**Route Diversion.** A motorist generally starts from a specific origin with an intended destination and desired route. The desired route is normally perceived to be the best route (based on some criterion such as historically experienced shortest travel time). However, this perception may not be accurate for the particular time and day in question. In fact, based on the simple criterion of travel time, the best route may change throughout the day and from day to day. Route diversion refers to the process of dynamically diverting traffic between alternative routes. This diversion should be determined in real-time by the traffic control system on the basis of current and preferably predicted traffic conditions. The recommended routes for a number of origin-destination pairs may be made available to motorists via information systems accessed by telephone, highway advisory radio, variable message signs or in-vehicle equipment such as vehicle/roadside communications (VRC) devices.
**Access Control.** Due to traffic congestion, traffic demand to certain facilities may require regulation. This may occur at entry points to freeways or toll plazas. *Ramp metering* and *pricing* are two common forms of access control. In the simplest form it may simply assume pre-time ramp meters to regulate traffic flow onto a mainline. In a more complicated form it may dynamically change ramp metering rates according to the arrival demand and the congestion on the mainline, and utilize methods such as real-time congestion pricing to regulate access to certain facilities.

**Mainline Control.** While access control involves controlling the access to any traffic facility, mainline control deals with controlling the traffic that is currently on the mainline. Mainline control must work in conjunction with access control and route diversion through devices such as variable message signs and vehicle/roadside communication devices. In addition, mainline control may be used by the traffic control system in real-time to regulate speed, lane usage, tunnel usage or controlling the movement of prohibited vehicles.

**Arterial Control.** The purpose of a traffic control system is to attain full utilization of all available facilities. This control function deals with the control of traffic on the local street networks and thus supplements mainline and ramp control. Arterial control consists of controlling traffic signals on frontage roads, parallel arterial and cross local streets.

Next section will present the proposed hierarchical feedback traffic control system that satisfies the desired properties (hierarchical, dynamic, closed-loop and proactive) and can be used for any of the freeway control functions reviewed above.
2.3 Proposed Hierarchical Feedback Traffic Control System

With the control system properties in mind, we next present a hierarchical feedback system for the dynamic traffic control system. The following requirements are imposed on the control system:

- Be functionally and/or geometrically distributed;
- Coordinate the control in the network
- Accept real-time surveillance data;
- Estimate and predict OD flows;
- Capture real-time traffic dynamics;
- Estimate current network conditions;
- Predict future traffic conditions;
- Optimize control strategies based on current and future conditions;
- Provide a stable control over all expected traffic conditions;
- Be able to run faster than real-time in order to implement on-line.

The structure of the system and its components are presented below.

2.3.1 Overall Structure

The structure of the proposed hierarchical feedback control system is shown in Figure 2.7. An inner feedback loop is employed for control management at the local level while an outer loop is used for the area-wide traffic management. This system consists of four modules: state estimation, OD prediction, local control and area-wide control. The state estimation module is to obtain the best estimates of network state given the available surveillance data. The OD prediction module takes the input from state estimation and predicts future origin-destination demand. Based on estimated state variables and predicted OD demand, the area-wide control module optimizes the control values
according to the overall system objective. The distributed local control module will then locally adjust the values set by the area-wide control to compensate for the exogenous disturbances and system errors (such as errors in state estimation and OD prediction). The inner loop is distributed over the network with little or no communication and computation effort while the outer loop is centralized for the entire network with intensive communication and computation burden. Consequently, the inner loop of the local feedback operates much faster than the outer loop of the area-wide optimization. For example, the time interval for the inner loop can be 10 to 30 seconds while the interval for the outer loop can range from 1 minute to 15 minutes.

2.3.2 State Estimation

As mentioned in previous sections, information obtained from the surveillance sensors can vary, depending on the type of the surveillance system employed. The state estimation module is used to process the surveillance data and estimate the current network state. The primary tasks for the state estimation module include:

1) the collection and aggregation of traffic data from various surveillance devices for a local region and for the entire area;
2) the detection of traffic accidents with their locations and severity;
3) the estimation of origin-destination demand;
4) the estimation of network state parameters such as densities, volumes, speeds and travel times.

The methodology for the last task (state parameter estimation) is primarily statistical and filtering techniques.
Figure 2.7 A Hierarchical Feedback System for the Dynamic Traffic Management
2.3.3 OD Prediction

Time-dependent OD flows are key inputs to the dynamic traffic control model. However, historical OD flows may not always give the actual travel demand in the network. The divergence of OD flows from their historical patterns can be caused by both supply and demand factors. An example of supply factors is the closure of roads or lanes that change the capacity of the network and an example of demand factors is the special events that temporarily increase volume to a particular destination. Both supply and demand factors can shift the equilibrium OD flows away from the historical observations. Consequently, one of the requirements for the dynamic traffic control modeling is the capability to predict OD flows in real-time.

2.3.4 Area-wide Control

The area-wide control module is a centralized predictive controller. The basic purpose of this module is to coordinate the operations of all the control devices. The coordination goal is achieved by the system-wide optimization in which each individual control parameter is synchronized to improve the overall system performance. The area-wide control module processes the information furnished by the OD prediction module, then predicts the future traffic state and generates the traffic control accordingly. Hence the control output is based on predicted data. Additional data required includes current traffic flows, speeds, queue lengths, the occurrence of incidents and the attributes of the network such as number of lanes, speed limit, length of each segment, etc. The control settings generated by the area-wide control module subsequently provide the nominal set values for the local control module.
Area-wide Control

Local Control

Figure 2.8 A rolling horizon approach to area-wide control
The area-wide control problem is generally modeled analytically as an optimization program. Therefore its solution is only optimal for a particular initial value. That is, the solution is open-loop. As mentioned previously, an open-loop solution is sensitive to exogenous disturbances and modeling errors. In order to overcome this shortcoming of open-loop, a rolling horizon approach is adopted in the area-wide control module. Figure 2.8 illustrates this rolling horizon method. First, the future traffic conditions are projected and the optimal control plans are solved for the entire projection horizon. Then only the control plan for the roll period is implemented to the local controllers. In the next run, the projection horizon is rolled one period ahead. A set of new control plans is then generated based on this new horizon and implemented for the roll period. This procedure is repeated for each run and the projection horizon is rolled ahead by one period each time with an overlap among different runs. As mentioned before, since the time interval for the local control is much shorter, the control plan given by the area controller will provide a set value for the local controller and the local control will do some fine-tuning to compensate the local traffic disturbance. Since the system conditions are updated at each run, this rolling horizon approach maintains a robust feedback control that has the capability to reduce the impacts of exogenous disturbances and prediction errors.

As will be reviewed in Chapter 3, existing models do not meet our architecture and modeling requirements, due to the following reasons:

- simplified representations of the transition of the system state (such as assuming traffic in steady-state or using turning fractions to model flow conservations);
- failure to couple the optimization and prediction,

A new dynamic analytical model is developed for this module. Features of this new model include:

- the model captures real-time traffic dynamics;
- the model maintains more accurate flow conservations by tracking OD flows instead of using turning fractions;
- the model optimizes control and predicts traffic simultaneously;
- the model has desired properties such as existence, uniqueness and consistence;
• the model can be solved efficiently in real-time.

A detailed mathematical formulation and solution algorithm will be presented in Chapter 4.

2.3.4 Local Control

In contrast to the area-wide control, the local control module consists of a set of distributed local controllers. Each controller functions independently and responds quickly to the changes in local traffic conditions. As mentioned before, the control values produced by the area-wide module provide nominal set values for the local controllers. In order to compensate traffic disturbances and prediction errors, some real-time adjustment is made by the local controllers based on local information. Many feedback control algorithms can be used in this control module. One such model that is well suited for local control is the linear-quadratic (LQ) feedback control model that will be reviewed in the next chapter.

2.4 Summary

A dynamic traffic control model has been integrated in this chapter. The following two approaches have been coupled in a hierarchical structure:

• long term, centralized, and predictive optimal control for area-wide control;
• short term, distributed, and reactive feedback control for local control.

Existing control algorithms for area-wide and local control will be reviewed in the next chapter and a new dynamic model for the area-wide control will be given in Chapter 4.
Chapter 3

Literature Review on Traffic Control Models

As mentioned earlier, the hierarchical, feedback approach to the design of traffic management system is not new. This chapter reviews major work in this area as reported in the literature. The first section reviews previous work in area-wide traffic control, followed by a review of existing local control systems. Next, some application studies are discussed. Finally, a brief comparison of various models and some observations from previous research are given.

3.1 Area-Wide Traffic Control

The optimal area-wide control problem was first formulated by Wattleworth and Berry in 1965 as a linear programming problem. Since then, many transportation researchers have discovered mathematical optimization techniques as a very useful tool for efficient control strategy designs. An area-wide traffic control model includes the following three main components:

a) A traffic flow model to represent the traffic process

b) Control variables

c) A performance index or objective function to be optimized
Existing area-wide control models differ from one another in their ways of modeling the traffic process, i.e., component a) above. They first can be classified as analytical or simulation-based models. The analytical models can be further divided into three categories: static, sequential and dynamic models. These three analytical models and a dynamic simulation model are presented below.

3.1.1 Static Control Model

The static optimal control strategies are derived from a priori known traffic data, such as average and historical demands. The static control model can produce fixed-time or time-of-day control policies, depending on the selection criteria.

![Diagram of freeway with on-ramp control](image)

Figure 3.1 A portion of freeway with on-ramp control

An early mathematical formulation of the static optimal control problem as a linear program was proposed by Wattleworth and Berry (1965) and Wattleworth (1967). As shown in Figure 3.1, a freeway corridor is divided into several homogenous sections. Each section is assumed to have one on-ramp and one off-ramp. The demand at each ramp and the origin-destination ratios are assumed to be fixed and known. The problem is to find admissible ramp metering rates that maximize the sum of input flows. Since the
traffic is assumed to be in steady-state, the objective of maximizing total in-flows is equivalent to maximizing total out-flows or throughput.

Mathematically, the problem can be formulated as the following linear program:

\[
\text{Maximize } J = \sum_{i=1}^{N} r_i
\]

\[
\text{Subject to}
\]

\[
\sum_{i=1}^{j} a_{ij} r_i \leq q_{j,\text{max}}, \quad j = 1, \ldots, N
\]

\[
0 \leq r_i \leq d_i, \quad i = 1, \ldots, N
\]

\[
r_{i,\text{min}} \leq r_i \leq r_{i,\text{max}}, \quad i = 1, \ldots, N
\]

where,

\( r_i \) = the \( i \)th on-ramp volume to the freeway or ramp metering rate (vehicle/hour);

\( r_{i,\text{min}} \) = the minimum metering rate for on-ramp \( i \);

\( r_{i,\text{max}} \) = the maximum metering rate for on-ramp \( i \);

\( a_{ij} \) = percentage of vehicles entering freeway at ramp \( i \) will pass through section \( j \);

\( d_i \) = demand at on-ramp \( i \) (vehicle/hour);

\( q_{j,\text{max}} \) = the maximum allowable volume or capacity on section \( j \);

\( N \) = total number of freeway sections.
Equation (3.1a) simply states that the objective function is to maximize the total input rates from all ramps. Notice that equation (3.1b) defines the freeway capacity constraints. Under the steady-state assumption, the freeway traffic volume $q_j$ in section $j$ is given by

$$q_j = \sum_{i=1}^{j} a_i r_i, \quad j = 1, \ldots, N$$

(3.2a)

Thus constraint (3.1b) is equivalent to the following freeway capacity constraint:

$$q_j \leq q_{j,\text{max}}, \quad j = 1, \ldots, N$$

(3.2b)

Constraint (3.1c) makes sure that the metering rate for any on-ramp can not either exceed the demand or be negative. The last equation bounds the metering rate between predicted minimum and maximum rates. The minimum metering rate must be less than the demand, $r_{i,\text{min}} < d_i$, otherwise the problem is infeasible.

This static control model is very simple and can be easily applied to any large-scale network. This model can also be solved efficiently by the well-known simplex algorithm. In addition, the model does not require any vehicle detectors on the freeway or ramps. In fact, most ramp meters implemented in the field use either fixed-time or time-of-day control.

There are two major disadvantages of this model. First, the control is not traffic-responsive, i.e., the ramp metering rates are independent of current traffic conditions. Second, the model assumes that the traffic flow is static or steady-state. The vehicle movement or flow propagation over the network is not considered. All vehicles are assumed to move at a constant speed in the network throughout the control periods.

### 3.1.2 Sequential Model

In order to overcome the non-responsive weakness of static control models, one approach has been to include real-time information about the traffic demand, e.g., the sequential
control model (Isaksen and Payne 1973, Papageorgiou 1980 & 1983, and E. Chang et al. 1994). Here the linear program in (3.1) is extended to multiple time periods as follows:

\[
\begin{align*}
\text{Maximize} \quad J &= \sum_{i=1}^{T} \sum_{t=1}^{N} r_{it} \\
\text{Subject to} \\
\sum_{i=1}^{j} a_{ij} r_{it} &\leq q_{j,\text{max}}, \quad j = 1, \ldots, N, \; t = 1, \ldots, T \\
0 &\leq r_{it} \leq d_{it} + \frac{lq_{it-1}}{\Delta t}, \quad i = 1, \ldots, N, \; t = 1, \ldots, T \\
r_{i,\text{min}} &\leq r_{it} \leq r_{i,\text{max}}, \quad i = 1, \ldots, N, \; t = 1, \ldots, T \\
lq_{it} &= lq_{it-1} + (d_{it} - r_{it}) \cdot \Delta t, \quad i = 1, \ldots, N, \; t = 1, \ldots, T \\
0 &\leq lq_{it} \leq lq_{i}^{\text{max}}, \quad i = 1, \ldots, N, \; t = 1, \ldots, T
\end{align*}
\]

where,

\begin{align*}
i &= \text{freeway section index;} \\
t &= \text{time period index;} \\
r_{it} &= \text{the ramp metering rate for on ramp } i \text{ at time } t; \\
d_{it} &= \text{demand at on-ramp } i \text{ at time } t; \\
lq_{it} &= \text{length of queue waiting at the entrance of on-ramp } i \text{ during time period } t; \\
lq_{i}^{\text{max}} &= \text{maximum length of queue allowed at ramp } i; \\
\Delta t &= \text{length of a time interval (hours).} \\
r_{i,\text{min}} &= \text{the minimum metering rate for on-ramp } i; \\
r_{i,\text{max}} &= \text{the maximum metering rate for on-ramp } i; \\
T &= \text{total number of time periods.}
\end{align*}
All other variables are defined as before. The objective function is the same as before except for the summation over $t$. Equations (3.3b) and (3.3d) are also the same as (3.1b) and (3.1d) in the previous model. The queue on the ramp is now measured by $l_{q_a}$. The real demand at the ramp $i$ during time period $t$ is given by the entering flow $d_{u}$ plus those vehicles waiting at the ramp, i.e. $d_{u} + l_{q_{i}}/\Delta t$. Thus equation (3.3c) guarantees that the metering rate can not be either greater than the demand or less than zero, which is analogous to (3.1c). Equation (3.3e) is new and it defines the ramp queuing dynamics over time. The term $(d_{u}-r_{u})\Delta t$ in equation (3.3e) is the excess demand during time interval $t$ and constraint (3.3d) limits the ramp queue to be less than a maximum value in order to prevent the spillback to surface streets.

Compared to the previous model, this sequential model considers time-dependent demands and explicitly models the ramp queuing evolution over time. This model requires two types of real-time input data: demand at each on-ramp during time $t$, $(d_{u})$, and the current number of vehicles waiting at the ramp, $(l_{q_{i}})$. These data can be provided by the surveillance system and an OD demand prediction model. Conceptually, this real-time information will improve the performance of the control model.

### 3.1.3 Dynamic Analytical Model

Although the sequential model includes demand and queuing dynamics, the representation of mainline traffic flow [equation (3.3b)] remains static since in each sequence (or time period) the mainline flow, $\sum_{i=1}^{j} a_{i} r_{i}$, is assumed to be in a steady-state condition. Consequently the traffic flow propagation from one section to another is ignored. Since frequent state transitions and rapid flow variations are observed in the most real traffic systems, especially during peak periods, the control strategies determined through steady-state models are of very limited use (Stephanedes and Chang 1992).
To overcome this shortcoming, various dynamic traffic control models have been
developed, some of which use a state feedback approach (Cremer and Fleischmann 1987,
and Zhang et al. 1995) and others which use a non-linear programming method
(Stephanedes and Chang 1993, and G.-L. Chang et al. 1994,). These dynamic models
differ from sequential models in their treatment of mainline traffic flows. These dynamic
models are based on the macroscopic traffic flow model (or LWR model) that were first
developed by Lighthill and Whitham (1955) and Richards (1956) as follows:

\[
\frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} = r - s \tag{3.4a}
\]

\[
q = f(k) \tag{3.4b}
\]

where \(q\) denotes the flow rate, \(k\) is the traffic density, \(r\) is the on-ramp flow, \(s\) is the off-
ramp flow, and \(x\) and \(t\) are space and time respectively. Equation (3.4a), which is the flow
conservation law, has a direct analogue in hydrodynamics. It basically states that the
change of concentration over time is equal to the in-flow minus the out-flow. The flow-
density relationship of equation (3.4b) is generally referred to as the fundamental diagram
of traffic flow (for a review of traffic flow theory, see May 1990).

A discretized version of the flow conservation law of (3.4a) is given by the following
equation:

\[
k_{it} = k_{i-1t} + \frac{\Delta t}{\Delta i} (q_{i-1it} - q_{it} + r_{it} - s_{it}) \tag{3.4c}
\]

where \(k_{it}\) and \(q_{it}\) are the density and the flow for section \(i\) at time \(t\), \(k_{i-1t}\) is the previous
density, \(q_{i-1t}\) is the upstream flow, \(\Delta t\) is the time interval length and \(\Delta i\) is the length of
section \(i\). Most existing dynamic control models use the discrete version of the flow
conservation law. The LWR model itself does not specify any functional form of the flow-
density diagram of (3.4b). Rather, various empirical functions have been derived by fitting
observed field data. Indeed, most existing dynamic models differ from one another in their use of this function. For example, Cremer and Fleischman (1987) and Stephanedes and Chang (1993) adopted a Greenshield type's model, G.-L. Chang et al. (1994) and Zhang et al. (1995) chose a triangular function, on which the flow is linearly increasing for light density and then linearly decreasing for density greater than the critical density.

Together with the traffic dynamics at on-ramps [which can be given by equation (3.3e)], the mainline traffic equations (3.4b) and (3.4c) constitute a complete dynamic traffic flow model. A dynamic optimal control problem can subsequently be formulated on the basis of this dynamic traffic flow model by solving the following:

\[
\text{Maximize } J(q,k,lq) \\
\text{Subject to} \\
\text{mainline traffic constraints (3.4b) and (3.4c)} \\
\text{capacity constraint } q \leq q_{\text{max}} \\
\text{ramp queuing constraint (3.3c-f)}
\]

where \( J(q,k,lq) \) is the objective function which depends on the state variables, flow \( q \), density \( k \) and queue length \( lq \). Various existing models differ from one another in their selections of the objective function, \( J() \), and the speed-density function of (3.4b).

Although in the literature most authors mention broad control objectives, such as the desire to reduce congestion, delays, travel times, pollution, and fuel consumption, the specific objective functions or performance indices vary. Generally speaking the following are common:

1) Minimize total travel time;
2) Minimize total delay;
3) Maximize total output volume (total flow leaving the freeway);
4) Maximize total input volume (total flow entering the freeway);
5) Minimize total number of vehicles in the queue;
6) Minimize carbon monoxide emissions and exposure;
7) Maximize total travel distance (single route case).

The resulting problem can be formulated as either a linear program (such as G.-L Chang 1994) or a non-linear one (such as Stephanedes and Chang 1993), depending on whether the traffic flow relationship is linearized or not. Theoretically, the optimal control problem can be solved by use of Pontryagin’s maximum principle (see Bryson and Ho 1975), however, the numerical calculations are intensive for a reasonable sized network. Instead, the problem is often solved by linear or non-linear programming algorithms.

3.1.4 Summary of the Analytical Control Models

The three types of area-wide analytical control models (static, sequential and dynamic) described above have many similarities. First, they all use a mathematical programming method to formulate and solve the system of equations. Second, all models optimize the overall control objective and thus the coordination of all on-ramps is implicitly considered. Finally, all use traffic demand information. The main difference among these models lies in their treatment of traffic flow.

In the static model, all traffic flows are assumed to be in steady-state. There are no time and space measures. In the sequential model, the ramp queuing phenomenon has been modeled but the mainline traffic flow is assumed steady-state. In the dynamic model, both mainline and ramp traffic flows are modeled dynamically.

3.1.5 Dynamic Simulation Model

Recently, a simulation-based dynamic model has been developed in the European DRIVE II project—called DYNA [Ben-Akiva et al. 1994]. The DYNA model can be illustrated by
Figure 3.2. It consists of two modules. One is a traffic simulator called the DYNAPredictor and the other is a control generator. The model first predicts future traffic conditions using the simulator, based on previous measurements and control inputs. Then the control generator produces the control inputs based on the prediction. Since the prediction depends on the control, the simulator is run again to forecast the traffic based on the new control inputs. This iterative procedure between prediction and control generation is conducted many times until consistency is reached.

The advantage of this approach is that it captures the traffic flow dynamics more accurately than an analytical model does and hence the simulation model may provide a
more reliable control solution. However the iterative procedure between prediction and control generation is a heuristic method and convergence for this procedure is not guaranteed. Even if the procedure does converge, it may give only a local optimum solution.

3.2 Local Traffic Control

In contrast to the area-wide control, a local control system considers only one isolated section of the network. Each local controller responds only to the changes in the local conditions. Figure 3.3 illustrates a local traffic control system for a freeway section. As shown in the graph, the traffic flow process for local control can be represented by the following measures:

- traffic volumes downstream and upstream: \( q_{out} \) and \( q_{in} \);
- traffic occupancy rates downstream and upstream: \( O_{out} \) and \( O_{in} \);
- on-ramp traffic volume (metering rate): \( r \);
- downstream bottleneck capacity: \( Cap \).

![Figure 3.3 A local traffic process](image-url)
Local control strategies can be classified into two categories. One is demand-capacity control and the other is linear quadratic control. Both are reviewed below.

3.2.1 Demand-Capacity Model

When traffic surveillance systems were first used on freeways in the 1970s, traffic engineers began to design local control strategies that utilize real-time surveillance information. The first of such designs is demand-capacity control, which is extensively used in the U.S. (Masher et al. 1975 and Koble et al. 1980). This local control is based on one measure, $q_{in}$, the traffic volume upstream of the merge area. The metering rate is determined by a real-time comparison of this upstream volume with the downstream capacity which is determined by historical data.

However, since the measurement of volume alone is insufficient to determine whether the freeway is congested or not, occupancy $O_{out}$ obtained from the downstream detector station is used for comparison. If the occupancy is above a preset threshold, congested flow is assumed to exist and the metering rate is kept at a minimum. Otherwise the metering rate equals the downstream capacity minus the upstream demand:

$$r = \begin{cases} 
    Cap - q_{in} & \text{if } O_{out} \leq O_{cr} \\
    r_{min} & \text{otherwise} 
\end{cases} \quad (3.5)$$

This control strategy is purely heuristic. Its philosophy is to maintain the downstream flow equal to capacity and to select a minimum metering rate when congestion exists. This model is not well suited for coordinated metering control and therefore is not suitable for implementation in a hierarchical architecture that contains both area-wide and local controllers.
3.2.2 Linear-Quadratic Model

Feedback is an important characteristic in automatic control theory. Researchers have used feedback control methods in traffic control since the 1970s [Isaksen and Payne (1973)]. The most popular local control strategy is derived from the linear quadratic (LQ) feedback control method, which has been intensively examined in the literature (Isaksen and Payne 1973, Athans et al. 1975, Goldstain and Kumar 1982, Papageorgiou et al. 1990 & 1991, Young 1994).

The LQ-regulator is a simple linear feedback control law that uses local measurements. The dynamic traffic process can be represented by:

\[ x_{t+1} = f(x_t, u_t) \]  \hspace{1cm} (3.6)

where \( x_t \) and \( u_t \) are state and control variables respectively. Equation (3.6) is also called transition equation, meaning that future state, \( x_{t+1} \), depends on current state and the control. Assuming nominal conditions about which to regulate the traffic, Payne (1971 & 1973) linearized the transition equation (3.6) around the nominal conditions and obtained the following linear equation

\[ x_{t+1} = Ax_t + Bu_t \]  \hspace{1cm} (3.7)

where \( A \) and \( B \) are constant coefficient matrices.

The objective of the local feedback control is to minimize the deviations from the nominal states \( (\bar{x}, \bar{u}) \). That is, the objective function is to minimize the following quadratic performance index:

\[ J = \sum_t \left\{ ||x_t - \bar{x}||^2 + w||u_t - \bar{u}||^2 \right\} \]  \hspace{1cm} (3.8)

51
where $w$ is a weighting factor. Minimization of this quadratic objective function subject to the linear traffic system of (3.7) leads to a feedback control law:

$$u_i = \bar{u} - L(\bar{x}, \bar{u})(x - \bar{x})$$  \hspace{1cm} (3.9)

where $L(\bar{x}, \bar{u})$ is the control gain matrix that can be calculated by solution of a Riccati equation (see Athans et al. 1975 and Papageorgiou et al. 1990 and 1991 and for detail). The resulting control law of (3.9) is called an LQ regulator. This feedback law is very simple and yet well-behaved in practice. If there is congestion (i.e., the measured state $x$ is higher than the desired value $\bar{x}$), the last term at the right hand side of (3.8) is negative and control is decreased relative to the nominal value. Similarly, if there is no congestion, the control input is increased above the nominal value.

### 3.3 Applications

Despite advances in the theoretical development of advanced traffic control systems, their implementation has been slow. For example, most existing ramp meters in the field today use either fixed control or heuristic based formulas such as demand-capacity control. Though many simulation studies have been conducted to test advanced freeway control strategies, most of them are simulated under a hypothetical network and traffic demand. However few field tests have occurred. For example, the tests in the I-94 freeway corridor in St. Paul, Minnesota (K. K. Chang 1990), the Boulevard Peripherique in Paris (Papageorgiou et al. 1991), the INFORM corridor at Long Island, NY (Gartner and Reiss 1987 and Reiss et al. 1991), the Garden Grove freeway at Orange County, California (Nsour et al. 1992), and the Amsterdam ring road network at the Netherlands (Young et al. 1994). Most of the simulation tests showed promising improvements over existing freeway methods. However, to date there still does not exist a comprehensive evaluation study to compare all proposed strategies.
3.4 Summary

Table 3.1 gives a summary of the various freeway traffic control models that have been reviewed above. For area-wide control, there are two main approaches: analytical and simulation-based. Analytical models use mathematical programming to formulate the optimal control problem. These models generally can be solved efficiently, especially for linear problems. However, they are not robust to modeling errors or disturbances in the traffic flow.

A simulation-based model for area-wide control has the potential to model traffic flows more accurately than any of the analytical models. Since a simulation-based model iterates between prediction and control generation, it is a closed-loop system. However, this simulation method lacks a rigorous theoretical foundation and many critical issues cannot be answered. These issues include:

- The convergence of the iterative procedure between prediction and control generation can not be guaranteed;
- Even if the procedure converges, the solution may not be globally optimal;

The analytical model offers an optimal solution for a coarse traffic flow model and the simulation-based model uses a fine traffic flow model but gives a solution that may not be optimal.

For local control, two types of models have been discussed. The demand-capacity model is a pre-timed heuristic model which has been widely implemented in the field. The linear-quadratic model is a feedback control model which has been intensively studied in the literature since 1970s.
<table>
<thead>
<tr>
<th>Control Range</th>
<th>Model</th>
<th>Control Logic</th>
<th>Methodology</th>
<th>Contribution Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area-wide Control</td>
<td>static model</td>
<td>pre-timed</td>
<td>linear programming</td>
<td>Wattleworth and Berry 1965, Wattleworth 1967</td>
</tr>
<tr>
<td></td>
<td>dynamic simulation model</td>
<td>dynamic, predictive</td>
<td>simulation</td>
<td>Ben-Akiva et al. 1994</td>
</tr>
<tr>
<td>Local Control</td>
<td>demand-capacity model</td>
<td>pre-timed</td>
<td>heuristics</td>
<td>Masher et al. 1975, Koble et al. 1980</td>
</tr>
</tbody>
</table>
3.5 Observations

Several observations should be pointed out from the literature review. First, the local control and the area-wide control must be combined together to form a hierarchical system. On the one hand, one centralized system requires intensive computation and communication linkage. Moreover, such a centralized system is often an open-loop system. On the other hand, a distributed local control system needs to coordinate all the local fields in order to overcome myopic behavior and achieve a network-wide objective.

Second, feedback control is a mature technique for a local controller. In particular, LQ feedback control is very easy to implement and it is robust in most situations.

Third, in contrast to the local control, area-wide control has not been fully developed. Existing area-wide models have many limitations. A good area-wide model for real-time applications should satisfy the following requirements:

- realistically represent traffic process over time and space;
- anticipate the effects of control measures on system performance;
- have the capability of control optimization;
- utilize real-time information such as origin-destination estimation and prediction;
- have desired theoretical properties such as existence and uniqueness; and
- be able to run faster than real-time.

Finally, comprehensive evaluation tests in both traffic simulators and in the field are needed. Some microscopic traffic simulators (for example, MITSIM at Yang 1993) are available for such evaluations. A simulation study is much less expensive and more flexible than a field test. Some standardized testing database that includes traffic networks and dynamic demand data should be established in order to create some standards to compare different strategies.
Chapter 4
Area-wide Dynamic Optimal Control Model

This chapter presents the primary contribution of this thesis. A new area-wide dynamic optimal control model is formulated and solved. This model is the heart of the hierarchical traffic control system that is presented in Chapter 2.

This chapter is organized as follows. Section 4.1 gives some background in traffic flow modeling. Section 4.2 presents the objective function of the model. Section 4.3 discusses in detail the mathematical formulation. Section 4.4 presents one solution algorithm and Section 4.5 gives concluding remarks.

4.1 Modeling Traffic Flow

Using conventional graph notations, a freeway network is presented by nodes and links. A node represents one particular location in the network, such as entry, exit, intersection or intersection of two sections. A link is the roadway that connects two nodes. Each link is further divided into segments. A segment is the smallest element in the network, within which the roadway is assumed to be homogenous in terms of network geometry (such as design speed, pavement condition, curvature, degree, number of lanes and lane width) and other characteristics (such as speed limit). Figure 4.1 illustrates an example of the freeway
network. It includes a mainline and several on-ramps and off-ramps. This sample network is redrawn in Figure 4.2 using a graph notation.

![Diagram of freeway network with on-ramps and off-ramps.]

Figure 4.1 An example of freeway network

![Network representation for the example freeway.]

Figure 4.2 Network representation for the example freeway

Vehicles on the freeway network are modeled as continuous flows. We will adopt the macroscopic traffic flow model (LWR model) developed concurrently by Lighthill and Whitham (1955), and Richards (1956). Though most of existing dynamic traffic control models also adopt this model, the dynamic flow conservation has not been enforced for a network with multiple origin-destination pairs. As reviewed in the previous chapter, existing analytical traffic control models (including both static and dynamic) use a priori
turning-fractions to model the flow conservation. This approach may be inaccurate for the
dynamic case in which the actual entry and exit flows deviate significantly from those
calculated by a priori turning-fractions. Moreover, the traffic flow in one segment is
represented by aggregate variables of density, flow and speed. Hence the vehicles cannot
be distinguished by their origin and destinations. Therefore, any OD-specific strategy,
such as route diversion, can not be implemented in such models.

To overcome these two shortcomings, the traffic flow is modeled here by each OD pair.
The dynamic flow conservation is enforced for each OD pair and thus eliminates the need
for the turning-fractions. Within each segment, vehicles can be identified by intended
destinations and thus route diversion strategies can be applied.

Let us define the following conventional notations:

- $N \subset Z_{+}$ set of nodes;
- $R \subset N$ set of origin nodes;
- $S \subset N$ set of destination nodes;
- $A \subset Z_{+}$ set of segments;
- $k_{i}^{rs}(t)$ density of traffic from origin $r \in R$ to destination $s \in S$ on segment $i \in A$ at
time $t$ (vehicle/mile/lane);
- $k_{i}(t) = \sum_{r \in R} \sum_{s \in S} k_{i}^{rs}(t)$ average density on segment $i \in A$ at time $t$ (vehicle/mile/lane);
- $f_{i}^{rs}(t)$ flow from origin $r \in R$ to destination $s \in S$ on segment $i \in A$ at time $t$
  (vehicle/hour);
- $\nu_{i}(t)$ average speed on segment $i \in A$ at time $t$ (mile/hour);
- $u^{r}(t)$ ramp metering rate at origin $r \in R$ at time $t$ (vehicle/hour);
- $d^{rs}(t)$ demand from origin $r \in R$ to destination $s \in S$ at time $t$ (vehicle/hour);
- $u^{r}(t)$ metered flow from origin $r \in R$ to destination $s \in S$ at time $t$ (vehicle/hour);
- $e^{r}(t)$ exit flow from origin $r \in R$ to destination $s \in S$ at time $t$ (vehicle/hour);
4.2 The Control Objective

The design of the traffic control system is guided by the control objective. As discussed in Chapter 2, there are many possible goals and objectives for a given control system. In this model, our goal is to improve the level of service and more specifically, our objective is to minimize the total vehicle travel time.

The total vehicle time spent on the mainline freeway during time period \([0, T]\) is given by

\[
T_1 = \int_0^T \sum_{i \in A} \sum_{r,s} [k_i^r(t) \cdot l_i \cdot n_i] dt
\]  

(4.1)

where \(\sum_{r,s} k_i^r(t) \cdot l_i \cdot n_i\) gives total number of vehicles on segment \(i\) during time \(t\). The total time spent in the on-ramp queues is given by:

\[
T_2 = \int_0^T \sum_{i \in A} \sum_{r,s} q_i^r(t) dt
\]  

(4.2)

where \(\sum_{r,s} q_i^r(t)\) is total number of vehicles waiting at segment \(i\) during time \(t\). The total vehicle travel time is the sum of \(T_1\) and \(T_2\):

\[
Z = T_1 + T_2 = \int_0^T \sum_{i \in A} \sum_{r,s} [k_i^r(t) \cdot l_i \cdot n_i + q_i^r(t)] dt
\]  

(4.3)

If time is discrete, the above objective can be rewritten as:
\[ Z = \sum_{i=0}^{r} \sum_{l \in A} \sum_{r,s} [k^r_i(t) \cdot l_i \cdot n_i + q_i^r(t)] \Delta t \]

(4.4)

where \( \Delta t \) is the length of each time interval.

### 4.3 Model Formulation

There are three types of constraints for the model: mainline traffic dynamics, ramp traffic dynamics, and capacity and non-negativity constraints. Each of them is discussed below.

#### 4.3.1 Mainline Traffic Dynamics

Consider the multiple origin-destination network flow problem. The number of vehicles on segment \( i \) from origin \( r \) to destination \( s \) at time \( t \) is \( k^r_i(t) \cdot l_i \cdot n_i \). The flow conservation law developed by Lighthill and Whitham (1955) and Richards (1956) is:

\[ \frac{d \text{(vehicles)}}{dt} = \text{in flow} - \text{out flow} \]

(4.5)

Denote in-flow as \( h \) and out-flow as \( g \) and we can apply the flow conservation law to the multiple OD network flow problem as follows:

\[ \frac{d[k^r_i(t) \cdot l_i \cdot n_i]}{dt} = h_i^r(t) - g_i^r(t) \quad \forall i, r, s, t \]

(4.6)

Figure 4.3 illustrates the flow variables for segment \( i \).
If the density on segment $i$ at an initial time $t = 0$ is equal to 0, i.e.,

$$k_i^r(0) = 0 \quad \forall i, r, s$$  \hspace{1cm} (4.7)

then the number of vehicles on segment $i$ at any time $t$ is given by

$$k_i^r(t) \cdot l_i \cdot n_i = \int_0^t [h_i^r(t) - g_i^r(t)]dt$$  \hspace{1cm} (4.8)

Given that the number of vehicles from origin $r$ to destination $s$ on segment $i$ at time $t$ is $k_i^{rs}(t)\cdot l_i \cdot n_i$, the number of such vehicles at time $t+\Delta t$ is given by

$$k_i^{rs}(t+\Delta t) \cdot l_i \cdot n_i = k_i^{rs}(t) \cdot l_i \cdot n_i + \int_0^{\Delta t} [h_i^{rs}(t) - g_i^{rs}(t)]dt \quad \forall i, r, s, t$$  \hspace{1cm} (4.9)

or

$$k_i^r(t+\Delta t) = k_i^r(t) + \frac{1}{l_i \cdot n_i} \int_0^{\Delta t} [h_i^r(t) - g_i^r(t)]dt \quad \forall i, r, s, t$$  \hspace{1cm} (4.10)

The discrete version of above flow conservation is given by

$$k_i^r(t+1) = k_i^r(t) + \frac{\Delta t}{l_i \cdot n_i} [h_i^r(t) - g_i^r(t)] \quad \forall i, r, s, t$$  \hspace{1cm} (4.11)

The total density on a segment is the summation of all the OD density on that segment:
\[ k_i(t) = \sum_r \sum_{i-1} k_{i-1}^r(t) \quad \forall i, t \] (4.12)

The entry flow from origin \( r \) to destination \( s \) at segment \( i \) can only come either from the upstream mainline

\[ h_i^r(t) = f_{i-1}^r(t) \quad \forall i \not\in R(r), r, s, t \] (4.13)

or from an on-ramp

\[ h_i^r(t) = u^r(t) \quad \forall i \in R(r), r, s, t \] (4.14)

Similarly, the exit flow from origin \( r \) to destination \( s \) at segment \( i \) can only go either to the downstream mainline

\[ g_i^r(t) = f_i^r(t) \quad \forall i \not\in S(s), r, s, t \] (4.15)

or to an off-ramp

\[ g_i^r(t) = e^r(t) \quad \forall i \in S(s), r, s, t \] (4.16)

The second type of constraints is the speed-density function.

\[ v_i(t) = V[k_i(t)] \quad \forall i, t \] (4.17)

Clearly speed must be a decreasing function of density, i.e., \( \frac{dV(k)}{dk} < 0 \). There have been many studies on the speed-density function such as Greenshields (1934), Greenberg (1959), Underwood (1961), Drake et al (1967) and Drew (1968). The Highway Capacity Manual does not give explicitly a single function for the speed-density relationship. Instead it gives an empirical curve for this relationship. Nevertheless, the following generic function seems to fit this empirical curve quite well (see May 1990):

\[ v_i(t) = V[k_i(t)] = v_f \cdot \left\{ 1 - \left[ \frac{k_i(t)}{k_{\text{jam}}} \right]^\alpha \right\}^\beta \] (4.18)

Here \( v_f \) is the free flow speed and \( k_{\text{jam}} \) is the jam density. \( \alpha \) and \( \beta \) are two model parameters. The combination of parameters \( \alpha \) and \( \beta \) defines the shape of the speed-
density function. These parameters are estimated from measured traffic data. All the empirical functions such as Greenshields, Greenberg, Underwood, Drake et al and Drew are special cases of this generic function. The typical shape of this speed-density function is depicted in Figure 4.4. This generic function will be used for our model.

The last constraint for the mainline traffic dynamics concerns flow. The well-known relationship between traffic density, flow and speed is:

\[ f_i^{in}(t) = k_i^{in}(t) \cdot v_i(t) \cdot n_i \quad \forall i, r, s, t \]  

(4.19)

which is known as the fundamental diagram.

![Speed-Density Relationship](image)

Figure 4.4 Speed-Density Relationship
4.3.2 Ramp Traffic Dynamics

Analogous to the mainline traffic dynamics, there are also three types of constraints for ramp traffic dynamics. The first constraint is the following queuing conservation law,

\[
\frac{d[q^\alpha(t)]}{dt} = d^\alpha(t) - u^\alpha(t) \quad \forall r,s,t
\]  

(4.20)

Here \( d^\alpha(t) \) is the OD demand, and \( u^\alpha(t) \) is the OD flow that can enter the mainline. Thus \( d^\alpha(t) - u^\alpha(t) \) is the excess demand at time \( t \). This is analogous to the mainline flow conservation equation. From the ramp point of view, OD demand \( d^\alpha(t) \) is the entry flow and \( u^\alpha(t) \) is the exit flow. Figure 4.5 illustrates this conservation law.

![Figure 4.5 On-ramp Queuing Conservation](image)

Analogous to equation (4.11), the discrete queuing conservation is given by

\[
q^\alpha(t + 1) = q^\alpha(t) + [d^\alpha(t) - u^\alpha(t)] \cdot \Delta t \quad \forall r,s,t
\]  

(4.21)

The on-ramp OD flow \( u^\alpha(t) \) is given by:

\[
u^\alpha(t) = \frac{d^\alpha(t)\Delta t + q^\alpha(t)}{\sum_s [d^\alpha(t)\Delta t + q^\alpha(t)]} \cdot u'(t) \quad \forall r,s,t
\]  

(4.22)

This implies that the on-ramp flow for a particular OD is proportional to its demand. The metering rate \( u'(t) \) has to be less than or equal to the total flow arriving at the ramp:

\[
u'(t) \leq \sum_s [d^\alpha(t) + \frac{q^\alpha(t)}{\Delta t}] \quad \forall r,t
\]  

(4.23)
Equations (4.23) and (4.22) also enforce that the on-ramp flow for each OD pair is less than or equal to the arrival flow:

\[ u^\alpha(t) \leq d^\alpha(t) + \frac{q^\alpha(t)}{\Delta t} \quad \forall r, s, t \]  

(4.24)

Finally the off-ramp flow is given by the fundamental diagram:

\[ e^\alpha(t) = k^\alpha_i(t) \cdot v_i(t) \cdot n_i \quad \forall i \in S(s), r, s, t \]  

(4.25)

### 4.3.3 Capacity and Non-Negativity Constraints

There are several capacity and non-negativity constraints that must be enforced. The density at each segment must be less than the maximum jam density.

\[ k_i(t) \leq k_{\text{jam}} \quad \forall i, t \]  

(4.26)

In order to prevent the spillback to the local streets, the queue length at each on-ramp must also be constrained to be less than the maximum queue length allowed.

\[ q^r(t) \leq q_{\text{max}} \quad \forall r, t \]  

(4.27)

For practical reasons, the control variable--ramp metering rate is bounded between a maximum rate and a minimum rate:

\[ u_{\text{min}} \leq u^r(t) \leq u_{\text{max}} \quad \forall r, t \]  

(4.28)

Finally, all the variables must be non-negative.

### 4.3.4 Modeling Traffic Flow Dynamics in Vector Form

In order to better understand the traffic flow dynamics, we can rewrite the traffic flow model developed above in vector notation. The traffic state at a given time can be represented by density, flow, speed and vehicle queue length at that time:

\[ X(t) = [k^\alpha_i(t), f^\alpha_i(t), v_i(t), q^\alpha_i(t)] \forall r, s, i \]  

(4.29)

Control variables are the entry flows to the mainline or the ramp metering rates:

\[ U(t) = [u^r(t)] \forall r \]  

(4.30)

The input variables are the dynamic origin-destination demands:
\[ I(t) = [d^r(t) \forall r, s] \]  \hspace{1cm} (4.31)

The system state transition can then be written as:

\[ \dot{X}(t) = G[X(t-1), U(t), I(t)] \]  \hspace{1cm} (4.32)

or

\[ X(t) = X(t-1) + G[X(t-1), U(t), I(t)] \]  \hspace{1cm} (4.33)

This dynamic state evolution is illustrated in Figure 4.6.
4.3.5 The Optimal Control Model

The dynamic optimal control problem can be stated as follows:

**Dynamic Optimal Control Problem:** Given an initial condition for the state variables $X(t) = [k_i^n(t), f_i^n(t), v_i(t), q_i^n(t)] \forall r, s, i$, and given predicted origin-destination demand $I(t) = [d^n(t)] \forall r, s$, find the on-ramp volume trajectories for each ramp over time $U(t) = [u'(t)] \forall r$, which minimizes the total travel time subject to state transition constraint of $\dot{X}(t) = G[X(t - 1), U(t), I(t)]$ and other capacity and non-negativity constraints.

This dynamic optimal control problem can be formulated as a discrete time non-linear program as follows.

\[
\begin{align*}
\text{Min} \quad Z &= \sum_{t} \sum_{i} \sum_{r,s} [k_i^n(t) \cdot l_i \cdot n_i + q_i^n(t)] \Delta t \\
\text{s.t.} \quad k_i^n(t + 1) &= k_i^n(t) + \frac{\Delta t}{l_i \cdot n_i} [u_i^n(t) - f_i^n(t)] \quad \forall i \in R(r), r, s, t \quad (4.35) \\
q_i^n(t + 1) &= q_i^n(t) + [d_i^n(t) - u_i^n(t)] \cdot \Delta t \quad \forall r, s, t \quad (4.40) \\
k_i(t) &= \sum_{r \in R} \sum_{s \in S} k_i^n(t) \quad \forall i, t \quad (4.37) \\
v_i(t) &= V[k_i(t)] \quad \forall i, t \quad (4.38) \\
f_i^n(t) &= k_i^n(t) \cdot v_i(t) \cdot n_i \quad \forall i, r, s, t \quad (4.39) \\
u_i^n(t) &= \frac{d_i^n(t) \Delta t + q_i^n(t)}{\sum_i [d_i^n(t) \Delta t + q_i^n(t)]} \cdot u'(t) \quad \forall r, s, t \quad (4.41)
\end{align*}
\]
\[ u'(t) \leq \sum_{i} [d^r_i(t) + \frac{q^r_i(t)}{\Delta t}] \quad \forall r, t \]  
(4.42)

\[ k_i(t) \leq k_{jum} \quad \forall i, t \]  
(4.43)

\[ \sum_{i} q^r_i(t) \leq q'_{\max} \quad \forall r, t \]  
(4.44)

\[ u'_{\min} \leq u'(t) \leq u'_{\max} \quad \forall r, t \]  
(4.45)

\[ k_{i}^{\alpha}(t), f_{i}^{\alpha}(t) \geq 0 \quad \forall i, r, s, t \]  
(4.46)

\[ v_{i}(t) \geq 0 \quad \forall i, t \]  
(4.47)

\[ u''(t), q''(t) \geq 0 \quad \forall r, s, t \]  
(4.48)

### 4.4 Solution Algorithm

This section presents a solution algorithm for the dynamic optimal control problem described above. Since the problem is a constrained nonlinear program, many algorithms for solving unconstrained NLP problems such as steepest descent, conjugate gradient, or Quasi-Newton methods (e.g., BFGS algorithm) can be applied with the special consideration of constraints. The convex combination algorithm (Frank and Wolfe 1956) which maintains the feasibility of the solution at each step is chosen to solve the dynamic optimal control problem formulated in equations (4.34)-(4.48).

In order to speed convergence, the diagonalization or relaxation technique is used. As shown in the flow chart in Figure 4.7, the algorithm contains two loops, one inner and one outer. At each inner loop, the vehicle speed is fixed, and the Frank-Wolfe algorithm is applied to solve for the optimal ramp metering rate, density, flow, and queue length. After the inner loop has converged, the vehicle speed is then updated at the outer loop. The program will terminate when the outer loop has also converged. This procedure is summarized as follows.
Step 1. Initialization.

Find an initial feasible solution $X^{(0)}(t) = \{u^{(0)}(t), k^{(0)}(t), f^{(0)}(t), q^{(0)}(t)\}$, and set outer iteration counter $n = 0$.

Step 2. Diagonalization.

Find a new estimate of segment speed $\nu^{(n)}_i(t)$ according to equations (4.37) and (4.38). Set the inner iteration counter $m = 0$.

Step 2.1: Direction Finding.

Based on the estimated segment speed $\nu^{(n)}_i(t)$, find the optimal moving direction $Y^{(m)}(t) = \{u^{(m)}(t), k^{(m)}(t), f^{(m)}(t), q^{m0}(t)\}$ by solving the following linear subproblem:

$$\text{Min } \nabla Z(X^{(m)}) \cdot Y^{(m)}$$

subject to

constraints (4.35), (4.36), (4.39-4.46) and (4.48)

Step 2.2. Line Search.

Find the optimal step size $\lambda^{(m)}$ that solves the one-dimensional search problem.

Step 2.3. Solution Update.

$$X^{(m+1)} = X^{(m)} + \lambda^{(m)} Y^{(m)} \quad (4.50)$$

Step 2.4. Inner Convergence Test.

If $X^{(m+1)} = X^{(m)}$ go to step 2; otherwise go to step 2.1.

Step 4. Outer Convergence Test

If $\nu^{(n+1)}_i(t) \approx \nu^{(n)}_i(t)$, stop. The current solution is optimal. Otherwise set $n = n + 1$ and go to step 2.
Initialize $X^{(0)}$, $n=0$

Update vehicle speed $V^{(n)}$

Find moving direction, $Y^{(m)}$

Search step size $\lambda^{(m)}$

Update solution, $X^{(m+1)}=X^{(m)}+\lambda^{(m)}Y^{(m)}$

Inner iteration converge?

Yes

Outer Iteration Converge?

Yes

Stop

No

No
4.5 Summary

This chapter presents an analytical dynamic optimal control model that is suitable for area-wide traffic control. It can be applied to the freeway control problem for a corridor that consists of a mainline and multiple entry and exit ramps. Compared to existing models, it has the following advantages:

- The model captures the true traffic flow conservation by distinguishing the traffic flows by their origin-destination pairs. In contrast, all of existing models use a priori turning fractions to model the flow conservation. This differentiation of flows by OD pairs also allows the implementation of those control strategies that use OD information, such as route guidance.

- The model uses a generic non-linear speed-density function which is valid under most traffic conditions. Existing analytical models adopt a linearized triangle function (Zhang et al. 1995 and Chang et al. 1994) which is a rough approximation of the speed-density relationship in either light traffic or heavy congestion situations. It is well known that the speed-density relationship is highly non-linear when the density is in the vicinity of its critical value. Therefore a linear approximation can not capture the traffic flow process under these conditions.

- The model optimizes the control settings and predicts the future traffic conditions simultaneously.

- The resulting model formulation is a mathematical program with a linear objective function but with non-linear constraints. A solution algorithm that combines the diagonalization technique and Frank-Wolfe method has been developed to solve the resulting system of equations.

One limitation of this model is its requirement of perfect O-D demand prediction. Therefore its optimal solution is sensitive to the prediction errors. A rolling horizon approach, in which the measurement and prediction are updated continuously and the optimal solution is repeatedly solved for each horizon, may reduce the impact of the prediction errors.
In addition, the impact of the prediction errors can be further minimized by incorporating a local feedback control in which the control value is adjusted in the real-time to the changing local conditions. As discussed in Chapter 2, a hierarchical structure can be used to combined the area-wide and the local control.

A case study that uses the hierarchical structure and the rolling horizon approach to combine the area-wide dynamic optimal control model and the local linear-quadratic feedback control model is explored in the next chapter.
Chapter 5
Dynamic Traffic Control Evaluation: A Case Study

This chapter presents a case study for testing the proposed dynamic traffic control model using microscopic simulation. The selected network is a portion of Interstate 93 in the Central Artery/Tunnel network. The control system implemented is ramp metering control and the aim of this case study is to (i) demonstrate the dynamic traffic control framework and (ii) test the proposed control algorithm.

5.1 Evaluation Framework

There are two ways to evaluate a control system. One way is by operationally testing the system in the field and the other is by testing the system via simulation. Field testing is more reliable since it operates in real-world situations but is too expensive for systems that may require infrastructure changes. Moreover, a field test is less flexible because the control system can only be tested against the scenarios that actually occur. Consequently, the performance under other scenarios will still be unknown. Microscopic simulation offers an inexpensive and yet rich laboratory environment for evaluating any new traffic control system. Here a number of scenarios can be staged to test the robustness of the control system to different assumptions on demand, driver behavior and events that reduce capacity (for a detailed scenario analysis, see Hotz et al. 1994)
The evaluation framework is illustrated in Figure 5.1. As shown, dynamic traffic control and the traffic flow processes form a feedback loop. The traffic control logic first determines the signal timings of the control devices, which in turn affect the individual drivers’ behavior and the overall traffic flow performance. The traffic flow performance is measured via the various surveillance devices that are modeled in the network. Based on these performance measures, the traffic control system subsequently generates new control actions.

![Figure 5.1 Evaluation Framework for a Dynamic Traffic Control System](image)

The studies presented here use the microscopic traffic simulator (MITSIM) developed at MIT (Yang 1993). MITSIM models traffic flow, control and surveillance at the microscopic level (i.e., individual vehicles, control devices and sensors are explicitly modeled) and moves vehicles according to car-following and lane-changing models. It also models in detail how drivers respond to other vehicles, traffic signals, speed limits, lane use signs, incidents and toll booths. Driver behavior is incorporated by various behavioral parameters such as desire speed, headway, gap, impatience level in following a
vehicle, etc. This level of detail is necessary for an evaluation at the "lane" level. In our evaluation studies, MITSIM represents reality or the actual conditions under which the dynamic traffic control model will be implemented.

As explained previously the dynamic traffic control model implemented here consists of three modules. The first is the state estimation module, which estimates the current traffic state based on the data obtained from the surveillance devices in the MITSIM. This output of this module includes current traffic speeds, flows, density, travel time, and vehicle queues, etc. The second module is a local feedback control module that provides a fast but limited control response to local measurements. The third module is the area control module, which optimizes the future control settings based on measurement of the current state in order to improve an area-wide objective. These optimized settings provide the nominal set values for the local controllers that are distributed across the network. Figure 5.2 illustrates the proposed control model and shows how the evaluation framework is implemented.
Figure 5.2 Implementation of the Dynamic Traffic Control Evaluation
5.2 Network and Case Study Design

This section gives the case study network and discusses the case study design.

5.2.1 Overview of The CA/T Project

The Central Artery/Third Harbor Tunnel (CA/T) project is a 7.5 mile interstate highway project that will replace the aged elevated Central Artery (I-93) with an enlarged subsurface expressway and build a new tunnel beneath the inner Boston harbor to improve access to the Logan International Airport. Approximately half of the 7.5-mile roadway is tunnel. When completed, it will be the largest and most complex underground highway system in the world. It will also be one of the busiest highways, carrying an estimated 250,000 vehicles daily through downtown Boston by the year 2010 and an estimated 100,000 vehicles per day through the new harbor tunnel. The project is designed to double the existing traffic capacity. Figure 5.3 shows the Central Artery/Third Harbor Tunnel network. In total, the CA/T network will comprise 107 lane-miles of roadway, 37 lane-miles of which will be covered sections or submerged-tube tunnels.

5.2.2 Case Study Design

The network used for this case study is a part of the overall Central Artery/Tunnel network consisting of south bound Interstate 93 and is shown in Figure 5.4. The mainline is a 2.7-mile freeway with 4 to 5 lanes in one direction. There are 6 on-ramps and 5 off-ramps. The network consists of 23 links, 70 lanes, and 24 origin-destination pairs. The detailed network representation is given in Figure 5.5.
Figure 5.3 The Central Artery/Ted Williams Tunnel project network
Figure 5.4 Case Study Network
Figure 5.5 Network Representation for I-93
Ramp metering control is the only control function examined in this case study. There is one ramp meter for each on-ramp in the network. The surveillance devices are loop detectors located on the freeway. The objective is to use ramp metering to improve the overall level of service in the network, including both the mainline and the ramps.

The simulation period is 60 minutes and the initial traffic demand on the network is low. The following scenarios with different traffic demand levels are simulated in this experiment:

1) Low demand, shown in Figure 5.6.
2) Moderate demand, shown in Figure 5.7.
3) High demand, shown in Figure 5.8.

The following four control algorithms are tested in this study:

1) No control;
2) A local control only;
3) An area control only;
4) A bilevel control that combines both local and area controls.

The three controls 2), 3) and 4) are all dynamic and adjust the metering rates in real-time. The detailed algorithms implemented for these controls are given below.
Figure 5.6 Low Demand
Figure 5.7  Moderate Demand
Figure 5.8 High Demand
A linear-quadratic (LQ) feedback control algorithm (see Section 3.3.2) is implemented for
the local control. The philosophy is to detect the mainline downstream occupancy of the
on-ramp and use this measurement to adjust the metering rate around a preset target
value. When this LQ algorithm is implemented alone (i.e., control algorithm 2), the preset
value is given by the previous metering rate. That is, the metering rate, $u_t$, at time $t$ is
given by:

$$u_t = u_{t-1} - L \cdot (o_t - \sigma)$$  \hspace{1cm} (5.1)

where $L$ is a constant control gain matrix; $o_t$ is current downstream occupancy; $\sigma$ is a
constant desired occupancy and $u_{t-1}$ is the previous metering rate. The time interval for the
local control is 60 seconds. This metering rate gives the green/cycle ratio for the signal
timing.

The area-wide dynamic optimal control algorithm developed in Chapter 4 is implemented
for the area control. It is coded in MATLAB\textsuperscript{\textregistered} script, and treated as sub-component of
the DTC (see Figure 5.2). It is launched by the DTC main program (coded in C++) as an
optimization engine. Once the dynamic optimal control problem is solved, it returns the
solution to the main program. Then the DTC will either use the optimal solution to
calculate the signal timings directly (for area control only) or pass this solution as
parameters to the local control (for bilevel control). The time interval for the area-wide
optimization process is 5 minutes.

In the bilevel control case, the local LQ algorithm is modified to coordinate with the area-
wide control. The preset values in the algorithm are given by the optimal solution
calculated by the area-wide control loop. That is, the metering rate is given by:

$$u_t = \bar{u}_k - L \cdot (o_t - \sigma_k)$$  \hspace{1cm} (5.2)

where $\bar{u}_k$ and $\sigma_k$ are the optimal metering rate and occupancy calculated by the area-wide
module at horizon $k$ (time interval $k$ for the area-wide control loop). Notice that the
desired occupancy in equation (5.1) is a constant in the entire simulation period, while here it may change in each horizon.

5.3 Results and Analysis

Each of the four control strategies (no control, local control, area control and bilevel control) are tested for three levels of demand. Therefore, there are a total of 12 scenarios tested in this case study. The following five measures of effectiveness (MOE) are used to evaluated the designs:

- density for each segment on the mainline at each time interval;
- cumulative number of vehicles arriving at destinations in each time interval;
- total number of vehicles arriving at destinations during the entire simulation;
- average travel time for each O-D pair;
- total travel time for all vehicles.

The first MOE measures the traffic distribution over time and space and thus gives the detailed information about how the traffic congestion propagates over the entire network. The second MOE measures the dynamic cumulative throughput over time and the third one measures the static total throughput. The fourth MOE is the average travel time from an origin to a destination. It includes the on-ramp delay and the mainline travel time. The last MOE is an aggregate measure of delay over all vehicles, including both the on-ramp waiting times and the mainline travel times.

The first three MOEs mainly reflect the traffic situations in the mainline but the last two MOEs of travel times reflect the traffic conditions in both the mainline and the ramps. These five MOEs are sufficient to evaluate the performance of the control algorithms. Although other MOEs can be collected, they are either closely correlated with above measurements or not applicable for the purpose of this evaluation. Each of above five MOE outputs for each scenario is examined below.
5.3.1 Density Profiles

Figures 5.9 through 5.12 give the density profiles for each control method when the demand is low. These figures show that there is little difference among the four control methods. Since the traffic is very light, there is no congestion and the control strategies are less effective. However, what is interesting is that none of the control strategies degrade the traffic performance.

Figures 5.13 through 5.16 show the results when the demand increases to the moderate level. The figures show that without metering control, congestion builds very quickly (in about 10 minutes). The density reaches as high as 140 vehicles/mile/lane, which is near the critical density on the freeway. Moreover, the jam density shows up in the entire downstream of the mainline from segment 10 to segment 22. More than half of the freeway is jammed for more than 30 minutes. With any of the metering methods, the congestion is almost eliminated. Although these density figures show that any of the control algorithms smoothes the traffic and eliminates the congestion, the bilevel algorithm shows the best performance.

Figures 5.17 through 5.20 show similar results when the demand becomes heavy. Without any control, the traffic congestion is quite severe. The traffic jam spreads further upstream to the entrance of the mainline. It also takes a longer time to clear the congestion. The local control algorithm eliminates most of the congestion but the densities downstream (segments 20 to 22) are still over the critical density (about 65 vehicles/mile/lane). Figure 5.18 also shows that the densities at segments 20 to 22 fluctuates from time to time around the critical density. This fluctuation occurs at the frequency of the local control loop (one minute). The density under the area control algorithm also fluctuates around the critical density but it does so at a lower frequency, as shown in Figure 5.19. This can be explained by the fact that the area control algorithm updates the metering rates at a longer time interval (five minutes). When the demand increases at a particular time, the control does not change immediately until the next time
interval. Still, the bilevel algorithm works best in this high demand case. Figure 5.20 shows that the traffic is smooth everywhere and there is no fluctuation during the entire simulation.
Figure 5.9 Density--No Control with Low Demand
Figure 5.10 Density--Local Control with Low Demand
Figure 5.11 Density--Area Control with Low Demand
Figure 5.12 Density—Bilevel Control with Low Demand
Figure 5.13 Density—No Control with Moderate Demand
Figure 5.14 Density—Local Control with Moderate Demand
Figure 5.15 Density—Area Control with Moderate Demand
Figure 5.16 Density—Bilevel Control with Moderate Demand
Figure 5.17 Density—No Control with High Demand
Figure 5.18 Density—Local Control with High Demand
Figure 5.19 Density—Area Control with High Demand
Figure 5.20 Density—Bilevel Control with High Demand
5.3.2 Cumulative Throughput

Figures 5.21 through 5.23 show the changes in cumulative throughput over time for each control strategy, compared to the no control case. When the demand is low (Figure 5.21), both the local and the bilevel controls improve the throughput. However, the area control slightly decreases the throughput for this case. When the demand changes to moderate (Figure 5.22), the bilevel control algorithm has the highest increase. The area control algorithm also performs well but the local control does not demonstrate much improvement. If the demand increases further to the high level (Figure 5.22), both the local and the area control algorithms have very close results, but the bilevel algorithm shows a significant improvement over the local and the area controls.

Overall, the bilevel control algorithm demonstrates consistent improvement over both the local and the area control algorithms, except in the low demand case in which the bilevel and the local algorithms give the same result.
Figure 5.21 Cumulative Throughput Increase vs. No Control -- Low Demand
Figure 5.23 Cumulative Throughput Increase vs. No Control -- High Demand
5.3.3 Aggregate Throughput

Table 5.1 summarizes the aggregate traffic throughputs for each scenario. The throughputs here are measured by the cumulative number of vehicles arriving at their destinations during the entire simulation period. The table also shows the percentage improvement in throughput for each algorithm over the no control case.

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Throughput (no. of vehicles)</th>
<th>Improvement (vs. no control)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No Control</td>
<td>Local Control</td>
</tr>
<tr>
<td>Low Demand</td>
<td>6659</td>
<td>6832</td>
</tr>
<tr>
<td>Moderate Demand</td>
<td>10981</td>
<td>11047</td>
</tr>
<tr>
<td>High Demand</td>
<td>11176</td>
<td>11730</td>
</tr>
</tbody>
</table>

From the table, we can see that metering controls generally improve the traffic throughput, except the area control algorithm in the low demand case. We also find that the throughput improvement under each control algorithm is higher when the demand level increases. This is rather intuitive. As demand increases, the uncontrolled traffic becomes more congested and thus throughput drops faster. As shown in the table, the local control and the area control performance is about the same. However, the bilevel control that combines both the local and the area control shows significant improvement over any
of the individual algorithms alone. For example, in the high demand case, the bilevel control increases the throughput over the no control case by 8.41% while the local and the area controls only improve by 4.96% and 5.15% respectively.

5.3.4 O-D Travel Time

The density profiles and throughput measurements in previous sections can only reflect the traffic conditions on the mainline freeway. While they all have shown improvement on the mainline, the conditions on the ramps have not been reflected. The O-D travel time, however, measures an average travel time from an origin to a destination. It includes the delay on the ramp and the travel time on the mainline. Therefore, the O-D travel time reflects the traffic conditions on both the mainline and the ramps.

In general, a ramp metering algorithm can reduce the mainline traffic by restricting the access on the ramps. However, if the algorithm achieves the mainline congestion relief at a high expense of the ramps, the traffic delays on the ramps may out-weigh the relief on the mainline. In such cases, the travel time might not necessarily decrease even if the density on the mainline drops. Therefore, travel time is an important MOE in this evaluation.

Tables 5.2 to 5.4 present the O-D travel time results for each scenario. To simplify the representation, only 7 major O-D pairs are listed here (there are total of 24 O-D pairs). These 7 O-D pairs have the same destination—the end of the mainline (node 23 in Figure 5.3). Each O-D pair here has the largest traffic volume among all the O-D pairs originating from the same node. Table 5.2 shows the O-D travel times for each control algorithm when the demand is low. The mainline travel time is the travel time from the entrance of the mainline to the end of the mainline (O-D 9-23). The travel time for ramp 1 is the travel time from ramp 1 to the end of the mainline (O-D 8-23) and likewise for other O-D travel times. The average mainline travel time under any of the control algorithms is slightly better but the ramp travel times are slightly worse than those without control.
Some of the increases in ramp travel times are due to the capacity losses because of the ramp signals. The area control shows that some ramps are slightly over-metered, but the differences among various control methods are not significant. It is clear that the three control algorithms do not degrade the overall system performance in this low demand case.

Table 5.3 shows the O-D travel time results when the demand is moderate. With metering controls, the travel times on the mainline and most of the ramp O-Ds drop significantly. For example, travel times on O-Ds 9-23 and 8-23 reduce more than 80 and 60 seconds respectively. However, travel times on O-D 14-23 increase significantly. As shown in the density profiles in Figures 5.13 to 5.15, segment 21, the downstream of ramp 6, is the bottleneck of the network. In order to reduce the congestion in this bottleneck, all three control algorithms meter ramp 6 heavily and this causes a delay on ramp 6. As seen in Table 5.3, local control responds to a congestion in the mainline by metering the immediate upstream on-ramps. Only two upstream on-ramps 5 and 6, are metered severely. The area and bilevel controls seem to balance the metering across all upstream ramps. The bilevel control clearly gives the best overall results.

Table 5.4 gives the results when the demand becomes heavy. The O-D travel time on the mainline (O-D 9-23) jumps to 379 seconds, up 130% from its 164 seconds in the low demand case. With any of the metering algorithms, this mainline travel time reduces to less than 200 seconds. This implies a drop of more than 180 seconds in the mainline travel time. However, the travel times for the ramps are mixed. Some of them increase and others decrease. Since O-D 9-23 (mainline) carries the largest traffic demand, a small reduction in its travel time contributes a large improvement in the system performance. As we will see in the next section, the overall system performance measured by total travel time is improved substantially under any of the control algorithms. While the local control meters heavily the ramps nearest to the congested location in favor of further upstream ramps, the bilevel algorithm shows a good balance across the ramps.
<table>
<thead>
<tr>
<th>Control methods</th>
<th>O-D pairs (origin node-destination node)</th>
<th>Average travel time (seconds)</th>
<th>Improvement over no control (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>none</td>
<td>ramp 1 (8-23)</td>
<td>149.96</td>
<td>-0.50</td>
</tr>
<tr>
<td></td>
<td>ramp 2 (0-23)</td>
<td>98.63</td>
<td>-1.68</td>
</tr>
<tr>
<td></td>
<td>ramp 3 (13-23)</td>
<td>86.31</td>
<td>-1.44</td>
</tr>
<tr>
<td></td>
<td>ramp 4 (10-23)</td>
<td>70.87</td>
<td>-1.70</td>
</tr>
<tr>
<td></td>
<td>ramp 5 (16-23)</td>
<td>59.78</td>
<td>-2.15</td>
</tr>
<tr>
<td></td>
<td>ramp 6 (14-23)</td>
<td>18.53</td>
<td>-0.81</td>
</tr>
<tr>
<td>local</td>
<td>area</td>
<td>163.02</td>
<td>-0.75</td>
</tr>
<tr>
<td></td>
<td>area</td>
<td>144.27</td>
<td>-1.68</td>
</tr>
<tr>
<td></td>
<td>area</td>
<td>103.70</td>
<td>-3.28</td>
</tr>
<tr>
<td></td>
<td>area</td>
<td>88.47</td>
<td>-1.07</td>
</tr>
<tr>
<td></td>
<td>area</td>
<td>71.97</td>
<td>-1.10</td>
</tr>
<tr>
<td>bivevel</td>
<td>area</td>
<td>150.77</td>
<td>-1.70</td>
</tr>
<tr>
<td></td>
<td>area</td>
<td>99.71</td>
<td>-1.70</td>
</tr>
<tr>
<td></td>
<td>area</td>
<td>61.47</td>
<td>-1.70</td>
</tr>
<tr>
<td></td>
<td>area</td>
<td>19.36</td>
<td>-1.70</td>
</tr>
</tbody>
</table>
Table 5.3 O-D Travel Times -- Moderate Demand

<table>
<thead>
<tr>
<th>Control methods</th>
<th>O-D pairs (origin node-destination node)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mainline (9-23)</td>
</tr>
<tr>
<td>none</td>
<td>270.96</td>
</tr>
<tr>
<td>local</td>
<td>189.86</td>
</tr>
<tr>
<td>area</td>
<td>186.93</td>
</tr>
<tr>
<td>bilevel</td>
<td>188.48</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>improvement over no control (seconds)</th>
<th>local</th>
<th>81.10</th>
<th>71.94</th>
<th>78.78</th>
<th>83.86</th>
<th>37.65</th>
<th>-59.73</th>
<th>-181.47</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>area</td>
<td>84.03</td>
<td>68.66</td>
<td>-17.77</td>
<td>-81.38</td>
<td>31.21</td>
<td>22.59</td>
<td>-189.03</td>
</tr>
<tr>
<td></td>
<td>bilevel</td>
<td>82.48</td>
<td>76.69</td>
<td>81.98</td>
<td>-7.42</td>
<td>29.90</td>
<td>22.43</td>
<td>-174.63</td>
</tr>
<tr>
<td>Control methods</td>
<td>O-D pairs (origin node-destination node)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>mainline (9-23)</td>
<td>ramp 1 (8-23)</td>
<td>ramp 2 (0-23)</td>
<td>ramp 3 (13-23)</td>
<td>ramp 4 (10-23)</td>
<td>ramp 5 (16-23)</td>
<td>ramp 6 (14-23)</td>
<td></td>
</tr>
<tr>
<td>none</td>
<td>378.87</td>
<td>325.15</td>
<td>259.95</td>
<td>346.17</td>
<td>313.83</td>
<td>119.70</td>
<td>150.29</td>
<td></td>
</tr>
<tr>
<td>local</td>
<td>198.54</td>
<td>182.43</td>
<td>129.78</td>
<td>259.27</td>
<td>169.09</td>
<td>257.90</td>
<td>557.07</td>
<td></td>
</tr>
<tr>
<td>area</td>
<td>192.72</td>
<td>318.95</td>
<td>409.93</td>
<td>586.02</td>
<td>354.80</td>
<td>234.98</td>
<td>429.75</td>
<td></td>
</tr>
<tr>
<td>bilevel</td>
<td>197.41</td>
<td>245.84</td>
<td>146.83</td>
<td>324.25</td>
<td>286.49</td>
<td>199.76</td>
<td>430.22</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Improvement over no control (seconds)</th>
<th>local</th>
<th>ramp 1 (8-23)</th>
<th>ramp 2 (0-23)</th>
<th>ramp 3 (13-23)</th>
<th>ramp 4 (10-23)</th>
<th>ramp 5 (16-23)</th>
<th>ramp 6 (14-23)</th>
</tr>
</thead>
<tbody>
<tr>
<td>local</td>
<td>180.32</td>
<td>142.72</td>
<td>130.17</td>
<td>86.90</td>
<td>144.74</td>
<td>-138.20</td>
<td>-406.78</td>
</tr>
<tr>
<td>area</td>
<td>186.15</td>
<td>6.20</td>
<td>-149.98</td>
<td>-239.84</td>
<td>-40.96</td>
<td>-115.28</td>
<td>-279.46</td>
</tr>
<tr>
<td>bilevel</td>
<td>181.46</td>
<td>79.31</td>
<td>113.12</td>
<td>21.92</td>
<td>27.35</td>
<td>-80.06</td>
<td>-279.93</td>
</tr>
</tbody>
</table>
5.3.5 Total Travel Time

The last MOE used in this evaluation is the total travel time. It is the summation of each vehicle's travel time, which is defined as the time required to traverse the network from its origin to its destination. Like the previous O-D travel time, the total travel time also reflects the traffic conditions in the mainline and the ramps as well. Table 5.5 shows the results. Again, the control algorithms show insignificant savings or even a slight increase in the travel time when the traffic demand is low. When the demand increases, all control algorithms show significant reductions in total travel time and the bilevel algorithm has the best performance among all algorithms.

Interestingly, the improvement in travel time is greater than that in throughput when applying the ramp metering. For example, in the moderate demand case, the improvement in travel time is 12.2% while the improvement in throughput is only 4.84%. This can be explained by the fact that the traffic speed increases more quickly than the traffic volume does when the congestion reduces.

From all the MOE outputs, it is evident that the bilevel control algorithm outperforms other strategies in all measurements under medium and high volume demands. However, no substantial improvements are achieved with any of the metering algorithms when the traffic volume is low.
Table 5.5 Simulation Results of Total Travel Time

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Travel Time (vehicle hours)</th>
<th>Improvement (vs. no control)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No Control</td>
<td>Local Control</td>
</tr>
<tr>
<td>Low Demand</td>
<td>146.6</td>
<td>139.6</td>
</tr>
<tr>
<td>Moderate Demand</td>
<td>474.0</td>
<td>442.2</td>
</tr>
<tr>
<td>High Demand</td>
<td>679.9</td>
<td>616.3</td>
</tr>
</tbody>
</table>

5.4 Summary

This case study has developed a simulation framework for testing the dynamic traffic control (DTC) system. Under this framework, a microscopic simulator and a dynamic traffic control module are integrated. The microscopic simulator represents the real world traffic, on which various candidate control algorithms in the DTC module are tested. The case study in the Interstate I93 network demonstrated the following important results:

The local LQ feedback control algorithm and the area dynamic optimal control algorithm show very close results, but the bilevel algorithm that combines the local and the area algorithms in a hierarchical structure shows a substantial improvement over any of the other strategies.
In conclusion, the results show that the bilevel algorithm developed in this thesis is a promising control strategy for the real-time freeway operations. Nevertheless, there are limitations of this algorithm, some of which will be addressed in Chapter 6.
Chapter 6
Conclusions and Future Research

This final chapter summarizes the thesis and gives some future directions for the development of the dynamic traffic control model.

6.1 Conclusions

The objective of this thesis is (i) to develop a dynamic traffic control model that is well suited for the real-time freeway operations; and (ii) to apply the proposed model for freeway ramp metering control using simulation.

Towards this end, we first propose a hierarchical feedback system for the dynamic traffic control. The system consists of four important components—state estimation, O-D prediction, local control and area-wide control. This system combines both feedback control (local control module) and predictive optimal control (area-wide control module).

We then develop a new model for the area-wide control problem. This area-wide traffic coordination problem is formulated as a dynamic optimal control model. Mathematically, it is a program with a linear objective function but non-linear constraints. Compared to existing algorithms, it has the following advantages:
• It captures the true traffic flow conservation by distinguishing the traffic flows according to their origin-destination pairs. The differentiation of flows by O-D pairs also allows the implementation of those control strategies that are specific to the O-D information, such as route guidance.

• It uses a generic non-linear speed-density function that is more applicable to the dynamic traffic environment in which the steady-state condition may not be maintained. It is well known that the speed-density relationship is highly non-linear when the density is in the vicinity of its critical value.

• It optimizes the control and predicts the traffic simultaneously.

• It can be applied on-line in a rolling horizon approach.

Finally, we test the proposed model and some of the algorithms for a freeway corridor control with ramp metering in a microscopic simulation environment. The testing results show that the bilevel algorithm which combines the local feedback control and the area-wide dynamic optimal control works best in all scenarios.

6.2 Future Research

The dynamic control model and its algorithm presented in this thesis has shown desired theoretical properties and demonstrated promising experimental results. However, in order to implement such a model in the field, further research is needed. Some of the issues that should be explored in the future include:

• More evaluation tests. Although preliminary laboratory testing results are promising, more evaluation tests on both the simulator and the field are needed. Further simulation tests in other scenarios and/or other networks will help us understand the robustness of the model. For example, the model relies on the real-time data collected by the surveillance sensors. It would be interesting to know how well the model will perform when some of the sensors fail or malfunction. Moreover, since the model
depends on perfect O-D prediction. The sensitivity to prediction errors should be investigated. An operational field test will also provide a real experience that any simulator can not substitute.

- **Efficient algorithms.** The area control module in the model needs to solve a non-linear optimization problem which can be a computational bottleneck for a large network. This can be an issue for the real time applications. Efficient algorithms that explores the structure of the area optimal control need to be developed.

- **Other control strategies.** There are a variety of advanced traffic control techniques available today. We need to apply the model to other control strategies such as speed control and lane control. Variable speed limit signs can be used to reduce the recurrent and non-recurrent queues and their shock wave effects in the freeways so that the traffic flow becomes smoother. Another example is to use lane control signs to influence lane selections and reduce the congestion due to merging and splitting.

- **Combining freeway and surface streets.** So far the model has been limited to freeway control. However, the freeway and surface street control systems should be integrated. There are two motivations for this integration. First, the impact of a seemingly isolated control system should be examined in a broader area. For example, if not carefully designed, a freeway control strategy would spill or divert some of the traffic to the surface streets. The improvement on the freeways should not be made at the expense of the local streets and vice versa. Therefore, the impact of any single control strategy should be studied in a broader network. Second, the control systems on freeway and local streets need to be coordinated in order to improve the overall system performance. Any of the isolated control strategies often renders a sub-optimal solution at a network-wide level.

- **Control and routing integration.** Recently there has been a lot of attention to dynamic traffic assignment or DTA. The traffic assignment aims at optimizing the
route selections given the traffic signal settings; while the traffic control aims at optimizing the signal settings given the route selections. Clearly these two traffic management efforts need to be integrated together to obtain a consistent and sustainable improvement.

- **Day-to-day impact.** Finally the long-term impacts of a traffic control strategy need to be considered. For example, a control strategy may shift original-destination demand patterns and route selections after drivers have some experience with the control system. Some of these long term effects can be studied using a day-to-day traffic dynamic setting that models the traffic pattern changes from day to day.
References


