A Methodology for the Design and Evaluation of Traffic Control Systems Using Microsimulation
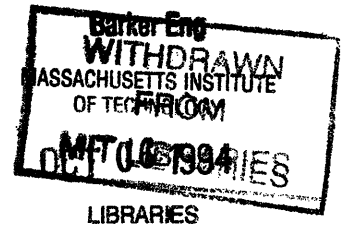
by

Nageswar Rao V. Jonnalagadda

Bachelor of Technology in Civil Engineering
Indian Institute of Technology, Madras, India
(1992)

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

August 1994

Signature of Author........
Department of Civil and Environmental Engineering
August 23, 1994

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

Certified by...............
Rabi G. Mishalani
Research Associate, Department of Civil and Environmental Engineering
Thesis Supervisor

Accepted by.......
Joseph M. Sussman
Chairman, Departmental Committee on Graduate Studies

A Methodology for the Design and Evaluation of Traffic Control Systems Using Microsimulation

by

Nageswar Rao V. Jonnalagadda

Submitted to the Department of Civil and Environmental Engineering
on August 23, 1994
in partial fulfillment of the requirements for the degree of
Master of Science in Transportation

Abstract

Most traditional traffic control systems consider few types of control and are static in nature. Advances in electronic and communication technologies, coupled with significant increases in computer computational power have made dynamic integrated traffic surveillance and control systems possible. A number of such control systems are under planning, operational, and testing stages all around the world. This thesis presents a simulation based methodology for the design, evaluation, and consequently design refinement of dynamic traffic control systems that integrate Advanced Traffic Management Systems (ATMS) and Advanced Traveler Information Systems (ATIS) aspects of Intelligent Vehicle-Highway Systems (IVHS).

The methodology focuses on gradually improving the design through iterative refinements. The design and evaluation process consists of four integrated components: (i) control and traffic flow simulator; (ii) traffic control system design; (iii) scenarios; and, (iv) performance measures. The simulator is microscopic in nature and is at the center of the whole process. It receives the control system design and scenarios as input. The traffic and control performance measures are obtained as output from the simulation.

For each of the objectives of the control system, control functions are designed as combinations of control devices and corresponding algorithms. Control systems are obtained as combinations of control functions catering to all the objectives. Scenarios capture other factors that influence traffic conditions and are specified by combinations of demand, events, and driver behavior. Traffic flow is simulated using the traffic control system design and the scenarios. The obtained performance measures are used to generate additional scenarios that further challenge the design. This scenario generation continues until the design is tested under a wide spectrum of scenarios. Moreover, the performance measures are analyzed to provide recommendations to refine the control system design. The design refinement continues until an effective, efficient, and robust design is achieved. Finally, a case study design for the demonstration of the methodology is presented.

Thesis Jointly Supervised by:

Moshe E. Ben-Akiva   Professor, Department of Civil and Environmental Engineering
Rabi G. Mishalani   Research Associate, Department of Civil and Environmental
    Engineering

Dedicated to my parents,

Venkateswarlu and Shakuntala Devi

Acknowledgments

I record with a deep sense of gratitude and appreciation the guidance provided by my advisors, Prof. Moshe Ben-Akiva and Dr. Rabi Mishalani. They have been a constant source of inspiration to me. Through the course of this research, I have learnt a lot from them. It has been a great opportunity to work with them and I feel very fortunate for this.

Many sincere thanks go to Tony Hotz for his valuable time and guidance. His valuable suggestions pulled me through many a problem. The innumerable prolonged discussions that I had with Tony and Rabi have made a huge contribution to this thesis. I have never met any one with such eternally jovial mood (is it because of an understanding of the essence of Gita?).

Thanks to my fellow researchers - Qi Yang, Kazi Ahmed, Rahul Herwadkar, and Owen Chen who have helped in their own way in different portions of the thesis.

I would also like to thank Bechtel/Parsons Brinckerhoff for their support of the Central Artery/Tunnel project at MIT, without which this research would not have been possible.

Thanks also to my friends at MIT, who are too many to list here, for their support at all times. They have helped in many ways to make me feel at home in this country.

My roommates Naresh, Rajeev, Arvind, and Prasad have made my stay in Cambridge an enjoyable experience. Thanks for their help, support, and friendship.

And finally, my family - parents, brother and sisters. I would not have been here had my parents not stressed on the importance of excellence. They have instilled in me a motivation and desire to succeed. I am eternally grateful to them for this. I especially appreciate the encouragement and inspiration my family has provided me from home.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

This thesis presents a methodology for the design and evaluation of real-time integrated traffic control systems using micro simulation models. This chapter motivates the research and describes the organization of the thesis.

1.1 The Worsening Congestion Problem

The usage of the roadways has increased enormously over the last two to three decades. The result has been the demand far exceeding the capacity, rendering the existing infrastructure inefficient. The problem has become predominant in the urban areas all around the world. All this has led to a high rate of congestion, resulting in long delays for individual drivers, large accident rates, reduced productivity, and enormous fuel wastage, and consequently environmental pollution.

Some statistics related to congestion reveal that in the United States, almost 70% of the peak-hour travel (vehicle hours) on urban Interstate highways and approximately 10% of all daily urban travel occurs under congested conditions [HS 90]. Also, the accident rates on a congested freeway are found to be about twice the rate on a freeway with freely flowing traffic [Kroes 83]. All this combined result in a loss of billions of dollars in production revenue annually. If left unchecked, the problem is expected to reach even higher levels in the coming years. For example, congestion delays are expected to increase by 400% by the year 2020 [Ben-Akiva 92] in the United States, without substantial improvements to the highway system. The need for exploring new ways of relieving traffic has never been greater.

1.2 New Approaches to Managing Congestion

Over the years, the standard solution to combat congestion has been the expansion of the transportation facilities - increasing the infrastructure resources; widening existing roads; and, building new expressways. Increasingly, expanding the facilities is becoming an infeasible approach due to the non-availability of sufficient space along with

environmental concerns. This has forced the search for alternate ways of controlling the flow of vehicles to avoid unnecessary congestion and reduce accident rates. This search gave birth to a new concept called Intelligent Vehicle/Highway Systems (IVHS) during the late 1980's. The basic idea involves taking advantage of the advances in electronic control and communication technologies to design traffic control systems that provide a higher level of performance by optimally utilizing the existing infrastructure without compensating safety, level of service, and the environment.

IVHS incorporates advances such as automatic traffic surveillance, incident detection and management, route guidance systems, fleet management, hazard warnings, electronic toll collection and congestion pricing, collision avoidance systems, in-vehicle driver aids and aims at eventually automating driving with full-time automated control. The activities of a fully deployed IVHS system are organized under six broad interrelated functional areas [Sussman 93]:

- **Advanced Traffic Management Systems (ATMS)**: ATMS integrates management of roadway functions. Real-time data is collected, utilized and disseminated by ATMS systems. It will predict congestion and respond in real-time to changing conditions by providing alternate routing instructions. Incident management and congestion pricing are some other critical functions falling under this area.

- **Advanced Traveler Information Systems (ATIS)**: ATIS provides information such as road conditions, weather problems, location of incidents, lane restrictions, optimal routing, and in-vehicle signing to travelers in their vehicles, homes, or places of work.

- **Advanced Vehicle Control Systems (AVCS)**: AVCS enhances the driver's control to make travel both safer and more efficient. This will include in-vehicle technologies like collision warning (alert the driver to a possible imminent collision) and avoidance (automatically break or steer away from a collision) systems, adaptive cruise control, fog and night-vision sensors, and radar. In the long-term, AVCS envisions vehicles running in closely spaced platoons, at normal highway speed, under automatic control.

The other three are major application areas that draw upon these three technological functions of IVHS.

16

- **Commercial Vehicle Operations (CVO)**: In CVO, the commercial vehicle operators adopt the ATIS and ATMS technologies to improve productivity of fleets and efficiency of their operations.

- **Advanced Public Transportation Systems (APTS)**: APTS uses the above technologies to enhance the accessibility of information to public transportation users. APTS also helps the public transportation operators to improve scheduling and the utilization of their fleets.

- **Advanced Rural Transportation Systems (ARTS)**: ARTS explores the possibility of economically applying the IVHS technologies in the rural areas where relatively low density roads are encountered.

IVHS envisions integrating all these advances to create a new transportation arena of "intelligent vehicles" on "intelligent highways". Under this IVHS umbrella, the solution to the traffic control problem requires new and innovative methods of information utilization and generation of signal controls. The integration of the three technical areas: ATMS, ATIS and AVCS provides improved prediction of traffic volumes and flows and better control of the traffic.

A number of projects incorporating these IVHS technologies are in the operational, testing, and planning stages around the world. Table 1-1 lists the various freeway surveillance and control systems in United States [IVHSMS 93]. In Europe, two major IVHS programs are underway: PROMETHEUS (Program for European Traffic with Highest Efficiency and Unprecedented Safety) and DRIVE (Dedicated Road Infrastructure for Vehicle Safety in Europe). Both projects address all the IVHS areas and deal with area-wide traffic control. Similarly, Japan is involved in a number of projects, including RACS (Road/Automobile Communication System), AMTICS (Advanced Mobile Traffic Information and Communication System), and VICS (Vehicle Information and Communication System) [SPIVHS 92].

Type of System	System Location
Areawide Systems:	
In Operation	Chicago Metropolitan Area Traffic Systems Center
	Denver Area Ramp Metering Control System
	Detroit SCANDI (Surveillance, Control ANd Driver Information) Systems
	Los Angeles Metropolitan Area Management System
	Maryland CHART System
	Minneapolis - St. Paul Twin City Traffic Management System
	San Diego Area Management System
	Seattle FLOW System
In Planning or Design	Boston Central Artery/Third Harbor Tunnel
	Columbus Metro Freeway Operations System
	Connecticut Statewide Traffic Management
	Dade County, FL FLAMINGO
	Fort Worth Traffic Management System
	New Jersey MAGIC (Metropolitan Area Guidance, Information, and Control)
	New York City Computerized Area Tracking System
	Orlando TRAVTEK
	Portland Freeway Management
	Phoenix Freeway Management System (Future)
	San Francisco Bay Area Traffic Operations Management System
Corridor Systems:	
In Operation	Covington, KY I-75 Reconstruction Projects Traffic Management Program
	Dallas North Central Expressway Corridor
	Long Island INFORM (Information for Motorists)
	Virginia I-66/I-395 Traffic Management System
In Design	Philadelphia I-95 Intermodal Mobility Project
	Santa Monica Freeway SMART Corridor Demonstration
Mini-Corridor Systems:	
In Operation	Delaware Wireless Emergency Phone System
	New Jersey Turnpike Automatic Traffic Surveillance and Control System
	New York Van Wyck Expressway Control System
In Planning or Design	Honolulu Trans-Koolau Tunnels System
Linear Systems:	
In Operation	Baltimore Jones Falls Expressway Surveillance and Control Project
	Cincinnati I-75 Traffic Diversion System
	Florida I-10/I-75 Motorist Aid System
	Florida Turnpike Motorist Aid System
	Los Angeles Artesia Freeway High-Occupancy Vehicle Commuter Lane System
	Massachusetts Motorist Aid Call Box Systems
	Orange County Costa Mesa Freeway High-Occupancy Vehicle Commuter Lane System
	Phoenix Black canyon Freeway Surveillance
	Seattle FLOW System
	Southeast Wyoming I-80 Motorist Information and Diversion System
	Wisconsin Milwaukee Freeway Control Project
Under Construction	Baltimore Area Surveillance and Route Diversion
In Planning or Design	Brooklyn Gowanus Expressway Contraflow Bus Lane Variable Message Sign System
	Newark Airport Interchange Surveillance and Control System
	Portland U.S. 26 Surveillance System
	Salt Lake City I-15 Corridor System

Table 1-1 U.S. Freeway Surveillance and Control Systems

Source: [IVHSMS 93]

18

1.3 Framework for an Integrated Traffic Control System

Research at MIT on IVHS includes the development of a Dynamic Traffic Management System (DTMS). An approach for evaluating such control systems has also been outlined, with specific reference to the Central Artery/Tunnel project of Boston [Hotz 94]. A fundamental characteristic of DTMSs is to react in real-time to traffic conditions. The generic structure of a DTMS represents a system with the following capabilities: (i) utilizes the real-time information provided by the surveillance system; (ii) predicts future traffic conditions; and, (iii) optimizes traffic control and routing strategies. The functional structure of the DTMS developed at MIT is illustrated in Figure 1-1 [Ben-Akiva 94].
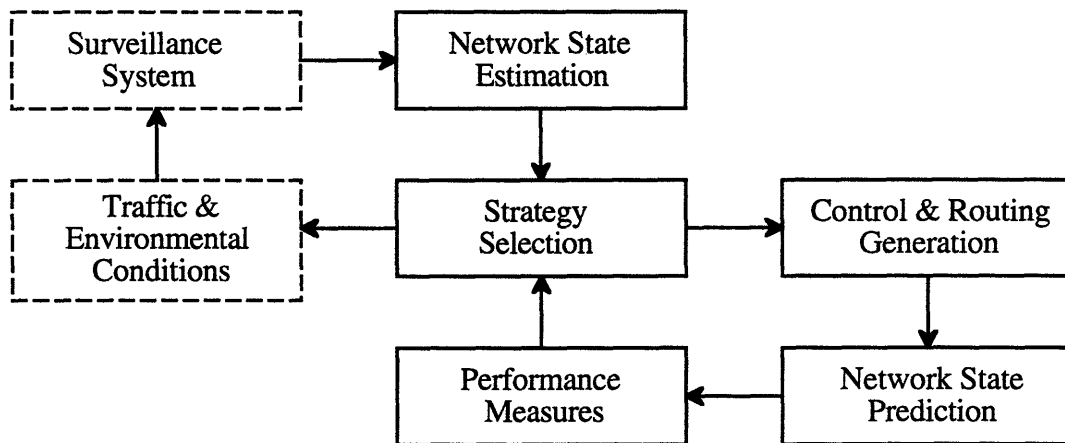
Figure 1-1 Dynamic Traffic Management System

The *Network State Estimator* estimates the current network state based on information gathered through the surveillance system. The *Strategy Selection* module serves two purposes. For a reactive system, this module represents a look-up table of network state versus control strategies. Based on the estimated state, the corresponding strategy is implemented. For a proactive system, the module evaluates alternative strategies based on predicted traffic conditions. This process proceeds as follows. The *Control and Routing Generation* module generates an initial control and route guidance strategy to alleviate congestion. Based on the current network states and the proposed control and

19

routing strategy, the future traffic conditions are forecasted by the *Network State Prediction* module. Corresponding performance measures are generated by the *Performance Measures* module. The *Strategy Selection* module then evaluates the strategy and analyzes its suitability to the traffic conditions. If found satisfactory, it is implemented on the network. Otherwise, additional strategies need to be generated thus invoking another generation-prediction-evaluation iteration. In these iterations, potential improvements are made to the control and routing strategy taking into account the predicted network conditions and the performance measures of the previously tested strategies. All these measures are used by the *Strategy Selection* module to evaluate various control and routing strategies generated.

1.4 Elements of the Integrated Traffic Control System

The integrated control system consists of three major elements: (i) Traffic and Environmental Surveillance; (ii) Motorist Information; and, (iii) Traffic Management.

1.4.1 Traffic and Environmental Surveillance

The function of the surveillance system is to detect the positions of the vehicles and pollution buildups in the network, thus providing the data about the traffic and environmental conditions. Various types of vehicle and environmental detection equipment are used to obtain this information. The direct output from the sensors is analyzed to obtain information on the traffic congestion and incidents on the network along with hazardous or unhealthy environmental conditions. This information from the sensors is used for incident response purposes and to supply input data to traffic management and motorist information systems. Sensor technologies are classified under four groups according to the type of data they obtain [Yang 93]:
- Point data
- Point to point data
- Area data and
- Other data.

Point data sensors are the most commonly used surveillance tools. These include inductive loops, infrared detectors, magnetometer, fiber-optic pressure detectors, and other

sensors that collect data at a point. These are normally used to collect traffic volume, occupancy, and possibly speed, height and type of a vehicle.

Point to point data sensors are used to obtain data like travel time, headways and volumes between two points. Two point detectors placed at a pre-specified distance can be used as a point to point sensor. A recently developed point to point sensor technology is Vehicle to Roadside Communication (VRC), often referred to as Auto Vehicle Identification (AVI) technology.

Area data sensors refer to those that collect, transmit and analyze video and area-wide radar data on traffic conditions for a road segment or an intersection. Some cameras provide visual surveillance for the operator. Data such as speed, volume, and density can be obtained from these.

Other data collectors include certain probe vehicles, police patrols, cellular phone call-in, overheight vehicle detectors, Carbon Monoxide (CO) and hydrocarbon sensors.

1.4.2 Motorist Information

The objective of the motorist information systems is to provide drivers with real-time information regarding prevailing traffic conditions. Providing information to motorists a priori and alerting them to downstream operational problems essentially serves three functions [Cassidy 92]:

- **Increased safety:** Motorists, once informed of the downstream roadway problems, are likely to approach the problem locations with greater caution.
- **Improved operation:** During congested operations, information on alternate routes could be provided to the drivers.
- **Improved public image:** A motorist information system is evidence of a state highway agency's commitment to improving highway safety and performance.

Information could be provided to the drivers in several ways. These are broadly grouped under the following:

Pre-trip Information. Pre-trip information is provided to travelers on road conditions, transit schedules, and paratransit opportunities. This assists travelers in making decisions regarding departure time, mode choice, route, and even destination. For example, information on costs, trip times, and departure and arrival

times may make shared-ride, HOV, and transit travel more attractive diverting travelers from personal automobiles to public transit.

Variable Message Signs (VMS). This is an effective means of providing information on current and anticipated traffic conditions. As the name suggests, messages on these signs can be updated when appropriate. These can also be used to provide route guidance. This is the simplest of the ATIS schemes and is already being implemented and tested in some parts of the United States.

Area Wide Broadcast. In this, Highway Advisory Radio (HAR) is used in providing traffic information and route guidance to drivers. HAR broadcasts can alert motorists to operational problems and provide detour information. The advantage of this over VMS is that drivers can get more detailed information at any point in the network, rather than at predetermined locations. This can also be used to supplement the motorist information provided by the VMS systems.

Automatic Route Guidance Systems. In this system, each driver is equipped with in-vehicle units that communicate with the Traffic Management Center (TMC). The information flow in these systems can occur both ways: from the TMC to the driver and from the driver to the TMC. Such systems have the potential of providing the most detailed and relevant information to drivers.

This thesis deals with the design and evaluation of integrated control systems utilizing the ATIS and ATMS functional areas of IVHS.

1.4.3 Traffic Management

Other than monitoring and informing motorists of roadway problems, the advanced technologies could be used to manage corridor operations during both recurrent and non-recurrent congestion. Traffic control on the freeway in response to real-time traffic information could be achieved using ramp meters, Lane Use Signs (LUS), Variable Speed Limit Signs (VSLS), directional signs, warning signs, mainline meters, and portal signals (in case of tunnels). The generation of route guidance is part of the traffic management schemes. Arterial signal optimization could better facilitate route diversion and aid in distributing diverted vehicles uniformly over the network. Incident response (response and dispatching of emergency vehicles) and congestion pricing are other operations under traffic management.

1.5 Literature Review

Most of the traditional traffic control systems consider few types of controls and are static in nature (static speed limit signs, ramp metering, and signalized control at intersections) [TCSH 76]. However, advances in electronic control and communication technologies, coupled with significant increases in computer computational power and improvements in systems engineering and operations research methodologies, present an opportunity for significant improvements in traffic control systems. Modern communications systems allow for the utilization of more information for traffic control than used by most traditional traffic control systems. These technological advances enable the design of a better control mechanism to improve system performance. The challenge is to design traffic surveillance and control systems that incorporate all the types of control discussed in the previous section in an integrated manner utilizing the advances in technology.

Section 1.2 presented an outline of the global efforts in designing such systems. Whether in US, Europe, or Japan, the traffic control systems designed are evaluated through operational tests. According to the 'Strategic plan for IVHS in the US' [SPIVHS 92], operational testing is an indispensable step that bridges the gap between research and development and full-scale deployment and provides an opportunity to conduct tests in a real-world environment under "live" transportation conditions. The operational evaluation results of only one system has been reported in the literature [Smith 92]: the evaluation of INFORM (Information for Motorists), a corridor traffic management system designed for a 40-mi highway corridor on Long Island, New York. The various INFORM control elements include overall supervision by operators in a control center, traffic monitoring, ramp metering, VMSs, and traffic signals at intersections. The evaluation was conducted through comparisons of vehicle miles of travel, average speeds, ramp delays, and motorist perceptions and other congestion related measures. However, evaluation through operational tests is very expensive. Computer simulation models allow for testing IVHS systems before conducting operational tests. This allows for major design refinements prior to implementation, thus resulting in more effective operational testing.

Various control elements of the traffic control system have been tested on an independent basis. For example, ramp metering [Shepherd 93, Stephan 92-1, Stephan 92-2], route diversion [Hobieka 92, Kwon 92, Rilett 92, Stephan 89], and variable speed limit signs [Smulders 90]. However, very few instances of work related to evaluating integrated

23

traffic control systems through simulation have been reported in the literature [Gartner 92, Riess 91].

In [Gartner 92], Gartner and Hou describe the application of a statistical method (based on ratio estimation techniques) for the comparative evaluation of alternative traffic control strategies. The technique is used in providing a measure of the accuracy of the estimates of the performance measures that are obtained as a ratio of two different parameters (e.g., average travel time obtained by dividing the total travel time by the number of vehicles) in terms of confidence intervals. Although relevant, this does not constitute a comprehensive evaluation.

In [Reiss 91], simulation studies conducted to assess the performance of Integrated Motorist Information System (IMIS) on the Long Island freeway corridor are described. IMIS integrates traffic monitoring through detectors, incident detection, incident management, ramp metering, variable message signs, and signalized intersections. Nine scenarios were studied by varying the traffic demands and the locations and severities of incidents. All the performance measures chosen (vehicle miles traveled, travel time, congestion clearance time after incident is removed, and maximum queue length) are positively correlated with each other, i.e., improvement in one measure will improve another measure. The aim was only to assess the benefits of the design and therefore it does not explicitly address the design and design refinement processes within an evaluation framework. The complexity of integrated control systems necessitates an iterative design, evaluation, and consequently design refinement approach. There are no methods reported in the literature that explicitly address the integration of the design of real-time integrated traffic control systems within an evaluation framework. This is what is addressed in this thesis.

1.6 Objective of the Thesis

The objective of the thesis is to develop a simulation based methodology for the design and evaluation of a real-time integrated traffic control system that integrates the ATIS and ATMS concepts of IVHS. The procedure for the design of a comprehensive control system is outlined. Since the design is a complex process, the preliminary design needs to go through a number of iterations, during which it is evaluated and refined, before a final design of advanced traffic control strategies could be proposed. Therefore, a detailed

methodology is needed to evaluate the performance of such strategies and to refine them before their implementation.

1.7 Thesis Organization

This chapter highlighted the increasing problems of congestion, introduced the IVHS concepts and brought out the need for a method to design and evaluate a control system within a unified framework. This section gives a brief description of the remaining chapters of the thesis.

Chapter 2 presents the framework for the design and evaluation of traffic control systems. An overview of the design and evaluation methodology is given and the different components of the overall evaluation framework are identified. Chapter 3 describes the evaluation methodology proposed in the thesis in detail. The simulation needs for such a methodology are initially identified. An approach for the traffic control system design is presented. Finally, the issues of how the various traffic demands, event patterns and types of probabilistic driver behavior are generated and the manner in which output of the simulation is used in refining the designs are addressed in detail. In Chapter 4, the design of a case study to demonstrate the methodology is described. Chapter 5 provides the conclusions of this research and discusses how this research could be extended.

2 Evaluation Framework

Evaluation of any control system essentially serves three major purposes. First, evaluation offers an opportunity to identify flaws in the design and hence, areas of further study. Second, it helps in measuring the stability and robustness of the control system. Finally, evaluation provides information that allows for refining the original design of the control system.

In this chapter, a framework is proposed for evaluating integrated traffic control systems. The chapter is divided into four sections. In the first section, the importance of identifying the goals and objectives of a traffic control system is presented. Section 2.2 highlights the importance of simulation in an evaluation process and outlines a methodology for evaluating traffic control systems. In section 2.3, the four major components of the framework - simulation, traffic control, scenarios, and performance measures are discussed. Section 2.4 closes the discussion with a more detailed description of the evaluation process and a reference to the performance measures and how to use them in making recommendations for refining the existing designs.

The description of the evaluation framework is illustrated through an example. The example is introduced in section 2.1 and the same example is used with the necessary and relevant details as the discussion proceeds through the chapter.

2.1 Definition of Goals and Objectives of Integrated Traffic Control Systems

The design of an integrated traffic control system is a complex process (refer to Chapter 1 for a description of an integrated traffic control system). A first step in formulating the design process is to identify the goals and objectives of the traffic control system. "Goals" define what we want to achieve through the control system and provide a conceptual basis for the design. Different ways of achieving each goal are specified as "objectives". These objectives determine how the control strategies are designed. Since a number of control devices can be used to achieve an objective, they must be coordinated to ensure that they work together. The evaluation process also depends upon the definition of

goals and objectives. Alternative control systems are compared with regards to the level of achievement of the goals and objectives they are designed for, in order to judge their performance.

A brief description of the identified goals is given below. The objectives to achieve each goal are also listed. A detailed explanation of each of the goals along with the corresponding objectives is given in section 3.3.

- Improve level of service: The primary goal of the control system is to improve the level of service. The congestion has to be reduced to provide a smoother traffic flow. This allows vehicles to reach their destinations with fewer stops and delays. The objectives aimed at achieving this goal include: (i) reduce link travel time; (ii) reduce origin-destination travel time; and (iii) reduce the time spent by vehicles in the queues.

- Improve capacity utilization: The main reason behind congestion is the demand on links exceeding their capacity. One way of increasing the capacity of the roadway is by appropriately changing control devices. Also, the capacity of the network can be better utilized by shifting traffic from routes of inadequate capacity to routes with excess capacity. The objectives aimed at achieving this goal include: (i) increase throughput; and (ii) reduce the number of lane changes.

- Reduce Carbon Monoxide emissions and exposure: Harmful emissions from vehicles deteriorate the environment and also affect the drivers. The aim is to improve the air quality, thus reducing the impact on environment and travelers. The objectives aimed at achieving this goal include: (i) reduce link density; (ii) reduce the variance of travel speed over time; and (iii) increase ventilation fan speed.

- Improve safety conditions: No goals should be achieved at the expense of safety because unsafe conditions cause accidents resulting in deaths, injuries and property damage. Also, accidents make it more difficult to achieve other goals. The chances of collisions between vehicles can be reduced by informing the drivers about any impending hazardous situations. This increases the caution applied while driving and makes travel safer. The objectives aimed at achieving this goal include: (i) reduce the number of lane changes; and (ii) reduce the speed differential across vehicles on the same link.

27

- Maintain control stability: The dynamic nature of advanced traffic control systems results in an unstable oscillatory behavior between the control system and the traffic conditions. Increased frequency in the changes in signal states is an indicator of such a situation. Under this condition, the control system should aim at reducing the variability in the device states, thus avoiding unstable oscillations. The objectives aimed at achieving this goal include: (i) reduce frequency of signal state changes.

- Control special vehicle access: The best way to reduce the impact of an incident is to clear it off in the least time possible. The access of special vehicles such as police cars, ambulances, and tow trucks to critical locations has to be increased to aid in faster return to normal conditions. Similarly, prohibited vehicles such as overheight vehicles must be curtailed from entering specific locations. The objectives aimed at achieving this goal include: (i) reduce travel time of emergency vehicles from origin to destination; and (ii) restrict prohibited vehicles from specific locations.

As an example, a control system with a subset of the above identified goals is considered. In particular, the two goals considered are:
- Improve capacity utilization and
- Improve safety conditions.

Each goal is assumed to be achieved through one objective. The two objectives are:
- Increase throughput for improving capacity utilization: This involves increasing the number of vehicles passing through the network in a given interval of time. This improves the carrying capacity of the network and hence, better utilization of the existing facilities.
- Reduce number of lane changes for improving safety conditions. The weaving maneuvers involved in a lane change cause a disturbance in the traffic conditions. Hence, a reduction in the number of lane changes will result in safer travel conditions.

28

2.2 Overview of Evaluation Methodology and Use of Simulation

Control systems designed for large networks are complex and may have a large number of diverse and complex interactions between the different control actions. Because of the complexities involved, a control system should be thoroughly tested before implementation. Along with testing, evaluation and design refinement are also required. A test provides the data necessary for evaluation and the results of the evaluation provide recommendations for design refinement.

Figure 2-1 illustrates the design and evaluation process. This process may be applied to advanced traffic control systems as follows. Initially, a control system design is proposed. This design is tested using simulation. The simulation output is evaluated and the recommendations for design refinement are proposed. The process is then repeated. This iterative procedure is necessary due to the complexity of the relationship between adaptive control systems and traffic flows.
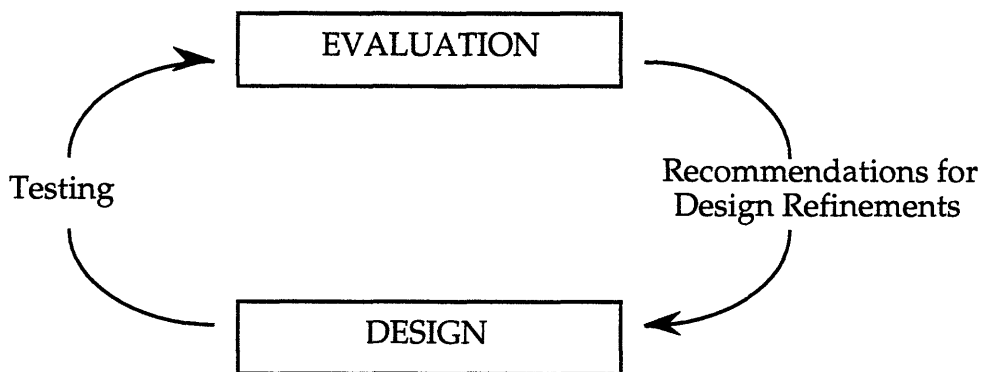
Figure 2-1 The Design and Evaluation Process

An advanced traffic control system is a *dynamic system* that measures traffic conditions and responds to these measurements by setting the states of traffic control devices such as LUSs, VSLSs, and VMSs, and controlling entrance ramp access and mainline flow. Moreover, congestion advisories, route guidance, and emergency information may be made available to motorists via VMSs, HAR, or special information provision on board vehicles. Such control actions alter the paths and the behavior of individual vehicles enroute to their destinations. In turn, the response of the motorists to

current control strategies impacts future control actions. This feedback or interaction between the control system and the traffic flow system is a key evaluation issue.

The large number of control devices, the different types of control logic that could be applied, and the different ways the devices can be used (independently or in combinations) allow for many possible control systems. The use of simulation to evaluate traffic control systems provides the opportunity to test many control systems. Moreover, simulation allows for the flexibility to change the functionality of different controls under different conditions and search for the best control combinations. For the simulation model to be useful for such an evaluation, it should couple the control logic with a microscopic traffic model that accurately represents the traffic flow.

For the simulation of all the details mentioned above, we need a simulator that captures the movement of each of the vehicles independently and simulates its behavior with respect to its mechanical and driver characteristics. The simulator should also realistically represent the operations of the surveillance system and the control system. This simulation is at the center of the whole evaluation process. Figure 2-2 illustrates the overall evaluation framework for integrated traffic control systems.
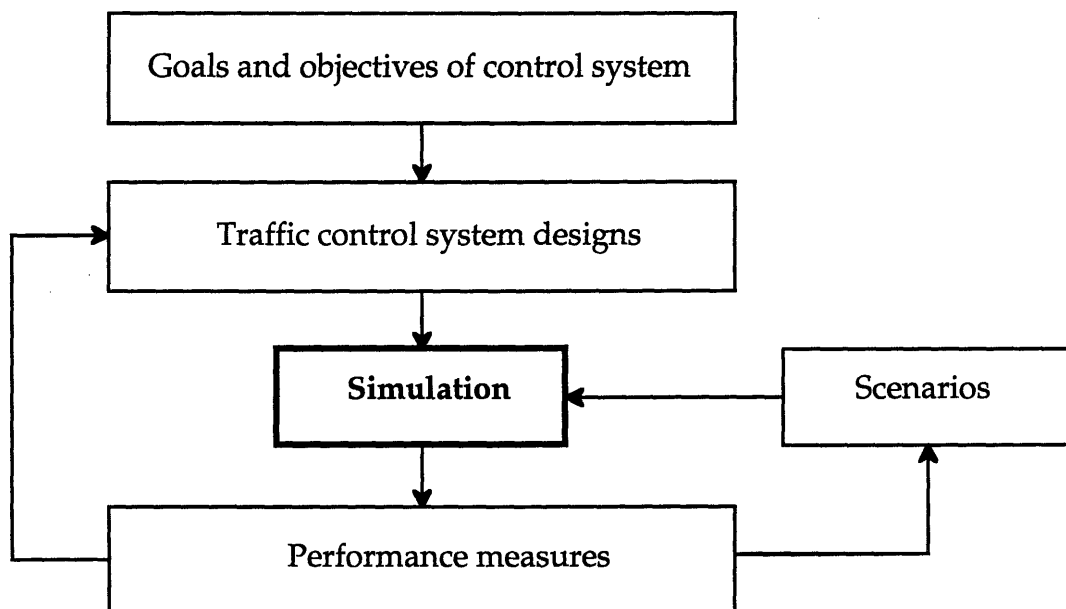
Figure 2-2   Overall Evaluation Framework for the Traffic Control System

30

As mentioned earlier, the design of the control system begins with the identification of the goals and objectives of the control system. The next step in the design process is the formulation of a traffic control system design that utilizes the available traffic control devices along with the algorithms in a coordinated manner to achieve the goals and objectives identified. This is one of the two major inputs to the simulation. The other major input to the simulation are 'scenarios'. Scenarios capture certain factors that affect the traffic conditions on the network. Specifically, these factors are - traffic demand, events (that capture the possible non-recurrent conditions on the network), and the varied behavior of the driver-vehicle pairs. A candidate control system is tested over a range of scenarios and performance measures are computed. These results are analyzed to suggest modifications to improve the tested design.

It is difficult to *a priori* determine the influence of the control system on the traffic conditions, hence the approach taken is to initially generate a control system design and execute the simulations, combining the design with a wide range of scenarios. Each design is tested by observing its behavior across the different scenarios. Based on the performance of the designs currently and previously tested, the design that performs best with respect to all the goals is selected. Further improvements are made to the selected design based on the deficiencies indicated by the evaluation. The new design is subjected to reevaluation. For example, if the measures indicate that the observed average speeds on two subsequent segments were drastically different during the same time interval, it means that more coordination of controls is needed. Similarly, a high queuing delay on the on-ramps with no delays on the mainline may be due to inappropriate ramp metering rates.

Similarly, since the exact effect of a scenario on traffic conditions is not known *a priori*, the formulated scenario may not cause the expected changes to the traffic conditions. The performance measures obtained from the simulation indicate the intensity of the effect of the scenario. This knowledge can be used to propose additional scenarios in order to further strain the control system and thus test its robustness. For example, a scenario may have resulted in high CO levels in the tunnel. This could be worsened by either increasing the traffic demand or by introducing a new incident inside the tunnel.

The four major components of the evaluation framework - simulation, traffic control designs, scenarios, and performance measures - are discussed in more detail in the next section.

2.3 Components of the Evaluation Framework

2.3.1 Simulation

As mentioned in section 2.2, the simulation is at the heart of the whole evaluation process. The simulator [Yang 93] is designed to model the performance of transportation networks under alternate traffic control systems. It takes the traffic control system design and the scenarios as input, and simulates the movement of vehicles in the network providing the performance measures as output. The main elements of the simulation model are:

- Network
  - Road network
  - Control devices
  - Surveillance devices
- Control logic
- Vehicle movement logic.

The integrated **road network** (freeways and urban streets) is represented in detail through its elements - nodes, links, segments and lanes. Nodes include external sources or sinks that receive and/or discharge vehicles, and internal nodes that lie between the source and sink nodes representing intersections. Links are directional roadways that connect nodes. It can be either a freeway, a ramp, or an urban street. A link is divided into geometrically homogenous segments recognizing the differences in the number of lanes, grade, horizontal curvature, or pavement conditions along its length. Each segment is further subdivided into its lanes.

The **devices** simulated for the **control system** include traffic signals, direction signs, warning signs, LUSs, VSLSs, VMSs, VRC devices, and HAR. These devices can be area specific (e.g., HAR), link specific (e.g., VSLS), or lane specific (e.g., LUS).

The operations of the **surveillance system** need to be realistically simulated to evaluate "real-world" traffic performance and control strategies. The various sensor technologies that are simulated include inductive loops, radar detectors, VRC devices, closed circuit television (CCTV), cellular phones, service patrol, and CO detectors. These devices are discussed in more detail in section 1.4.1.

The **control logic** is an integral part of the simulation. Based on the observed and predicted traffic conditions, it sets the states of the various control devices and provides route guidance information to relieve congestion. These decisions in turn affect the traffic conditions. Thus, the simulator couples the control logic with the traffic flow.

The most important part of the simulator is the movement of vehicles - which essentially includes the changes in vehicle positions, speeds, acceleration and deceleration rates and lane changes. This is captured in the simulator using models such as the car-following model and the lane changing model.

2.3.2 Traffic Control System Designs

Most traditional control systems are static in nature. That is, the same control logic is applied independent of the traffic conditions. For example, on freeways, the common control devices include directional and warning signs, static speed limit signs, and ramp meters. The only dynamic nature is due to the variation in device states, if any, during AM and PM periods. For example, the ramp meters may be preset to have different metering rates during the AM and PM periods. However, advanced traffic management systems will use real-time data from the network to set control devices such as LUSs, VMSs, and VSLSs. In such control systems, the traffic conditions are monitored via advanced surveillance equipment and control decisions are implemented in real-time. That is, the states of the control devices are changed based on the observed and predicted traffic conditions. Since the observed traffic conditions are obtained through the surveillance system, based on the data needs, the surveillance system design may need to be changed. Thus, a control system design includes the surveillance system design. However, the example presented focuses on traffic control design only.

The number of control devices and the different ways of using each device render considerable flexibility in developing control strategies. However, the interactions between the various control devices have to be kept in mind during the design process. The control logic will have to be designed to ensure that the devices work cooperatively and not against each other in achieving the desired objectives. For the same type of traffic conditions, the control devices and their logic can be combined in various ways to form a control system.

The different traffic control system designs are used as input to the simulation. Each design specifies a way in which a set of traffic control devices are used. The set itself may differ from one design to another.

With respect to the example discussed before, the control strategy for each of the objectives is designed using a set of control devices and a specific logic to control the device states. For example, for increasing throughput, ramp meters can be used to meter the traffic on the ramp based on the upstream demand and downstream capacity. Similarly, VMSs can be used to inform drivers not to change lanes frequently. Alternative control systems could be achieved by varying the set of control devices and the logic used in designing the controls for either or both the objectives. For example, the control system above could be altered by using a different logic to control the ramp meters or by using mainline meters instead of ramp meters to increase throughput.

2.3.3 Evaluation Scenarios

A wide variety of traffic conditions are possible on a network. A control system that performs well under certain conditions is not guaranteed to do so under all conditions. For example, a design that works best under uncongested conditions may breakdown under congested conditions. Therefore, the control system should be tested against a wide spectrum of scenarios to judge its robustness. A scenario along with the control system will influence the nature of the traffic conditions realized on the network.

A scenario is defined in terms of the following three factors: traffic demand, events and driver behavior parameters. Traffic demand dictates the volume of traffic on different parts of the network. Events impose possible non-recurrent conditions on the network. The behavioral parameters capture the probabilistic behavior of the drivers. These three factors combine in various ways to give different types of traffic conditions to be simulated on the network.

Each of the three factors can be varied over a wide range. Each combination of the three factors defines a scenario. Since a prohibitively high number of combinations are possible and since it is necessary to capture a wide variety of scenarios, a guided approach is needed for the generation of a set of scenarios for the simulation.

2.3.4 Performance Measures

Performance measures are obtained as outputs of the simulation. The performance measures are used for comparing alternative control system designs. The microscopic simulation allows the measures to be specified in as much level of detail as necessary. To be of use in the evaluation process, the measures should be chosen to exhibit the following characteristics:

- *Relevant to objectives*. Each measure should be clearly related to an objective defined in the design process. It is important that each measure used corresponds to one of the objectives, especially in problems with multiple objectives like the one on hand. The reasons will become clear in section 3.5.2 where attempts are made to utilize the measures to evaluate alternative control systems.
- *Manageable*. The number of measures used in the evaluation should be as small as possible, yet enough for a comprehensive evaluation.

The performance measures serve two major purposes. On one hand, the performance measures offer a very good basis for testing the performance of the control system. Based on the measures obtained, any deficiencies in the control system can be detected and appropriate design refinements can be made. On the other hand, the performance measures are indicative of the effect of the scenario. This in turn allows for synthesizing worst-case scenarios.

Different measures have been identified keeping all the above characteristics under consideration. The measures are outlined in section 3.5. A detailed specification is presented in Appendix A. Different measures that utilize the same basic output from the simulation are grouped under a common measure type. Table 2-1 lists out the measure types and the different performance categories (indicative of the goals) that they could be used to measure. The various measures under a measure type may capture various goals and hence, each type may embrace more than one category.

35

Table 2-1 Categories of performance measures

Measure Type	Performance Category (Goal)					
	Level of Service	Capacity Utilization	CO emiss & expos.	Safety Condition	Control Stability	Spec. Veh Access
Carbon Monoxide			√			
Travel Time	√	√				√
Speed	√			√	√	
Lane changing	√			√	√	
Density	√	√				
Queuing	√	√				
Freq of signal state changes					√	
Vehicle Miles Traveled		√				
Vehicle Hours Traveled		√				

2.4 The Evaluation Process

Section 2.2 provided an overview of the methodology for the evaluation of an integrated traffic control system. In this section, the evaluation methodology is explained in greater detail. Figure 2-3 shows the different steps involved in the evaluation process.

Initially, the goals of the control system and the objectives to achieve each goal are identified. The next step involves the identification of the relevant control functions that could potentially be used in achieving each of the objectives. A control function specifies the functionality of a set of control devices that work together to achieve a specific objective.

```
┌─────────────────────────────────────────┐
│   Goals and objectives of control system │
└─────────────────────────────────────────┘
                    │
                    ▼
┌─────────────────────────────────────────┐
│    Choice and design of control functions│
│           for each objective             │
└─────────────────────────────────────────┘
                    │
                    ▼
┌─────────────────────────────────────────┐
│     Selection of traffic control system  │
│             configurations               │
└─────────────────────────────────────────┘
                    │
                    ▼
┌────────────────────┐      ┌──────────────┐
│    Simulation      │◄─────│   Scenarios  │
└────────────────────┘      └──────────────┘
                    │
                    ▼
┌─────────────────────────────────────────┐
│          Performance Measures            │
└─────────────────────────────────────────┘
                    │
                    ▼
┌─────────────────────────────────────────┐
│  Recommendations for the control system  │
│            design refinement             │
└─────────────────────────────────────────┘
```
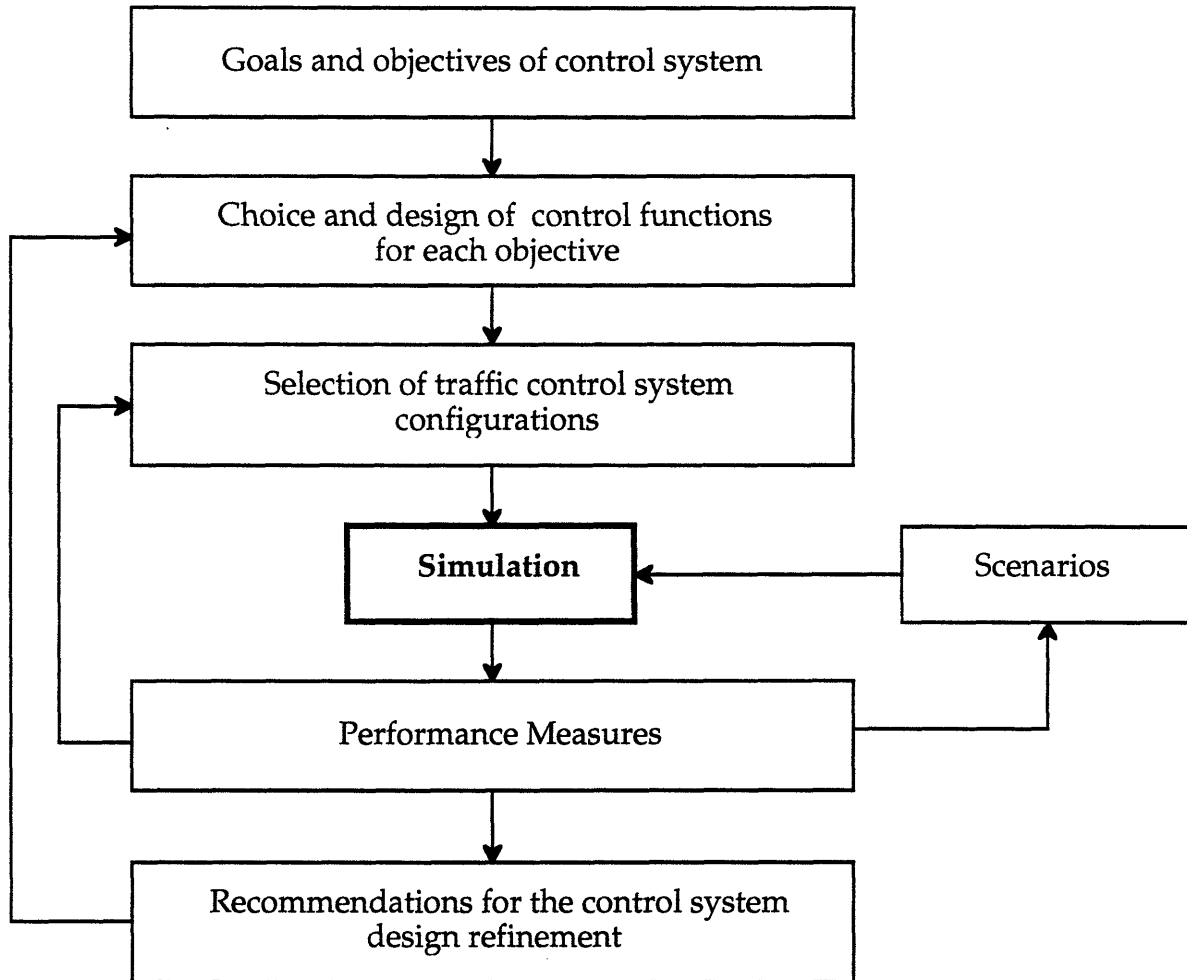
Figure 2-3 The Evaluation Process

For the example considered before, assume that there are three control functions and four different devices. The three control functions available for use in achieving the objectives are:

- Route Diversion: Diverting the vehicles to routes with excess capacity to relieve congestion.
- Mainline Control: Controlling the traffic on the main freeways.
- Access Control: Controlling the access of vehicles to the mainline.

The control devices available to design the control functions are:

- Ramp meters (RM)
- Mainline meters (MM)
- Variable speed limit signs (VSLS)

37

- In-vehicle information devices (IID)

As mentioned before in subsection 2.3.2, the focus of this example is only on the control devices and not on surveillance system design. The control devices that could potentially be used to design the control functions are shown in Table 2-2.

Table 2-2 Control devices for control functions

Control Function	RM	MM	VSLS	IID
Route Diversion				√
Mainline Control		√	√	
Access Control	√			

The control functions and control devices that could potentially be used in achieving each of the objectives are shown in Table 2-3.

Table 2-3 Control devices and control functions used for achieving objectives

Objective	Rte.Div	Mainline Control		Acc Con
	IID	MM	VSLS	RM
Increase throughput (Improve Capacity Util.)	√	√	√	√
Reduce no. of lane changes (Impr Safety Cond.)		√		√

The last step before performing any simulations is the identification of the traffic control system configurations. A combination of these control functions linked to specific objectives is referred to as a traffic control system *configuration*. A control function linked to an objective is hereafter referred to as a 'configuration component'. In the example considered above, assuming that each of the control elements have only one algorithm, a total of 5 configuration components are possible:

- Route diversion for increasing throughput
- Mainline control for increasing throughput
- Access control for increasing throughput
- Mainline control for reducing number of lane changes
- Access control for reducing number of lane changes.

A configuration can consist of a single component or a set of components. Hence, given the above components, 31 configurations are possible.

The control system design and refinement consists of two loops: the inner loop, formed by the configurations, the simulation, and the performance measures; and the outer loop, formed by the control functions, simulation, and the recommendations. While the inner loop deals with the testing of the various combinations of the components, the outer loop deals with the functional design of the various components. For each outer loop, the inner loop is visited a number of times. The details of each loop and how they communicate with each other is explained below.

Given the configuration components, the configurations for use in the simulation are enumerated. In the inner loop, each one of the configurations is used consecutively in the simulation. For each simulation, the performance measures are obtained. Initially, the configurations with single and two configuration components only are chosen and executed. Each of these configurations is simulated under a wide range of scenarios. The performance measures obtained in this iteration are used in deciding the configurations of higher number of components to be simulated in the following iterations. For example, if the results indicate that a combination of components reduces the benefits, that combination is not used subsequently. However, the main purpose of these inner loop iterations is to quantify the nature of the interactions between the various components. Such a quantification allows for refining the functional design of the components in terms of how they communicate with each other. In the final iteration, based on the feedback from the preceding iterations, combinations of control functions catering to all the objectives (control systems) are selected and tested against a common set of scenarios. This is to aid in the comparison of the performance of the alternative control systems.

Considering the example above, let A represent the objective of increasing throughput and B, the objective of reducing the number of lane changes. The combinations of components for the first iteration of inner loop are:
- Route Diversion for A
- Mainline Control for A
- Access Control for A
- Mainline Control for B
- Access Control for B
- and all the two-way combinations of the above components such as:
    Route Diversion for A + Mainline Control for B.

The other input variable to the simulation is the 'scenarios'. Each configuration chosen for simulation is combined with a wide range of scenarios to test for its robustness. To begin with, an initial scenario is generated for simulation. The performance measures are used to modify this scenario and create worst-case scenarios. These scenario generation iterations take place for a specific control configuration (i.e., for the first configuration simulated in an inner loop iteration). All configurations in an iteration use the same set of scenarios to serve as a basis for the comparison of their performance. Refer to section 3.4 for more details on the generation of scenarios for simulation.

Once all the relevant configurations in the inner loops are tested, the outer loop gets activated. The performance measures from the configuration tests are analyzed and recommendations are made regarding changes, if any, to be made to the algorithms used in the control function designs. These are used in generating a new set of configurations and the whole process is repeated until a final design is obtained. If necessary, control device locations could also be changed. Also, depending upon the data needs, changes could be made to the surveillance system. For more details, refer to section 3.3.

The performance measures as extracted from the simulation may not make much sense. Only when associated with the specific objectives that they reflect, can any interpretations be made. Also, in many cases, more than one measure is associated with an objective. In such a case, the measures have to be combined in some way to measure the performance of that objective. For example, assume that the performance of a control system with regards to 'safety' is judged based on the two measures: 'variance of speed of vehicles passing a point in a time interval' and 'variance of speed of a vehicle over its journey'. A higher speed variation at a point in a time interval means that the speeds of vehicles passing through that point during that time interval are widely different which makes travel unsafe. Similarly, a high standard deviation of speed of a vehicle over its journey means that the vehicle has been accelerating and decelerating frequently which affects safety conditions. If a design performs better or worse than another with regards to both measures, then, the decision as to which design is preferable can be made easily. On the other hand, if the design performs better on one measure and worse on another, then, the decision process becomes more complicated. More details on the approach taken to solve this are given in section 3.5.

In the iterative process of the outer loop, the analysis of the results may indicate that a configuration performs well with respect to some objectives and not so well with respect

to others. At this step, the aim of the design refinement process is to improve the performance of the control functions related to the objectives on which the performance was not upto requirements, while ensuring that the performance of the control functions related to other objectives is not traded off. Essentially, in every iteration of the outer loop, a set of alternative control system designs are chosen and tested under a wide range of scenarios to decide how to enhance their performance and bring them closer to a fully integrated control system.

3 Evaluation Methodology

3.1 Introduction

Chapter 2 presented a framework for evaluating real-time integrated traffic control systems. In section 2.2, the importance of using simulation to evaluate such control systems was presented. The different components of the network that need to be simulated and the need for a comprehensive microsimulation model is justified in the next section. In order to perform the evaluation and design refinement process (described in section 2.4 and illustrated by Figure 2-3), an initial set of alternate control systems should be specified. However, a large number of control systems are possible and this necessitates a systematic evaluation approach. The evaluation approach includes three basic steps:

- Formulating a process for generating alternate control systems
- Detailing the various scenarios that need to be simulated to test the robustness of the control system
- Using the performance measures obtained from the simulation to compare alternate designs of the control system and generate new designs.

The remainder of this chapter addresses each of the above steps in detail. Thus, this chapter deals with the four main components of the evaluation framework outlined in section 2.3.

3.2 Traffic and Surveillance and Control Simulation

A simulation model should have the following capabilities in order to be able to perform a proper evaluation of a traffic control system:

- support multiple vehicle classes
- support different driver characteristics
- support different levels of surveillance system
- support a wide range of traffic control strategies
- support different types of control devices and motorist information systems
- provide various measures of system performance.

In order to support the evaluation of integrated and dynamic traffic control and surveillance systems at the operational level, a detailed representation of traffic control operations coupled with traffic flow behavior is necessary. Also, control actions such as mainline control, incident detection and response, and emergency vehicle operations have detailed traffic input requirements. Microscopic simulation models are the only ones that provide this level of detail and the above capabilities. The microscopic simulator developed at MIT [Yang 93, Ben-Akiva 94, Hotz 94] satisfies the necessary requirements for such an evaluation.

To perform a proper evaluation of the system design, the microscopic simulator should replicate the physical characteristics of the network and the functional characteristics of the control system. The basic components that need to be simulated are:
- Traffic network geometry
- Traffic flow
- Surveillance system
- Control logic system
- Control device system.

The **traffic network** consists of multiple lane highways and tunnels, ramps, toll booths and the urban street network. The road network is normally represented in a hierarchical fashion and includes nodes, links, segments and lanes.

The **traffic** is made up of individual vehicles that have different characteristics such as maximum and desired speed profiles, acceleration and deceleration profiles, emission profiles, driver behavior, and traffic information availability. At the beginning of the simulation, each vehicle is assigned a path that leads to its destination. Subsequently, it is moved from one link to the next along its path. A vehicle's remaining path may be changed due to the driver's enroute decisions.

The **surveillance system** monitors the traffic conditions on the network. The surveillance system consists of: (a) a wide range of sensors and detectors such as loop detectors, Vehicle to Roadside Communication (VRC) devices, Closed Circuit Television (CCTV), overheight vehicle detectors, vehicle type detectors and CO sensors; (b) external agency information such as police reports and cellular phone call-ins; and (c) historical information stored in the control center.

This information on the real-time traffic conditions on the network is continuously processed and made available to other elements of the system that make control decisions. The historical information is used when the real-time information is not available (due to device failure) or is realized to be invalid (due to device malfunction). The **control logic system** predicts the travel times and congestion on the network and generates traffic control strategies.

The control decisions made in the control center are conveyed to the drivers and implemented on the network through the control device system. The **control device system** includes regulatory devices such as traffic signals, VSLSs, ramp meters, LUSs, and mainline meters; and advisory devices such as VMSs, HAR and in-vehicle systems. The changes applied to the control devices in turn influence the traffic flow conditions on the network.

The performance measures that are required for system evaluation, such as travel time, speed, density, time spent in queues, and percentage of vehicles crossing the CO exposure thresholds are obtained as outputs of the simulation model. These measures and their evolution over time can be output at a single location, lane, segment, link, OD pair, and entire network level.

3.3 Traffic Control System Design

The traffic control system is part of an overall traffic management system. Other functions of the traffic management system include operations such as processing the data obtained from the surveillance devices, maintaining communication between the data collectors and data processors, controlling the different facilities in the network and managing information. The traffic control system itself includes operations such as managing congestion, incidents, and special events and controlling overheight vehicles and hazardous vehicles.

This section provides a methodology for the design of a traffic control system and is divided into four subsections. As a first step, we need to identify what the goals and objectives of the control system are. A description of the goals and objectives is given in subsection 3.3.1. In transportation networks, the traffic conditions in one part of the network is dependent on that of another. Hence, every control action may affect other

control actions. Therefore, the importance of the interaction between the different objectives needs to be realized. These interactions are discussed in subsection 3.3.2. Subsection 3.3.3 examines the means of achieving the different objectives. The last subsection presents a systematic approach for generating various control system configurations for evaluation. These configurations help in designing a comprehensive control system.

3.3.1 Goals and Objectives

As was presented in section 2.1, six goals have been identified that should be considered in the design of any traffic control system. The goals are chosen to complement each other and to be as much independent of each other as possible. The six major goals are:

- Improve level of service
- Improve capacity utilization
- Reduce Carbon Monoxide emissions and exposure
- Improve safety conditions
- Control special vehicle access.
- Maintain control stability

These goals are achieved by using different methods referred to as 'objectives'. Even though objectives are specific to goals in most cases, the same objective may sometimes be used to achieve different goals. The following is a brief description of the goals, and the objectives under each goal that aid in its achievement.

Improve Level of Service

A transportation facility should accommodate the traffic demand with an acceptable quality of service. The level of service is defined in the traditional manner for our purposes. The indicators of level of service include travel speed, travel time, and traffic interruptions. Three objectives are used in achieving this goal.

- Reduce travel time on O-D pair:
    The travel time on an O-D pair is defined as the time taken by a vehicle leaving origin O to reach destination D. A number of links join together to form a route between an origin and a destination. An O-D pair may potentially have multiple routes. In case of O-D pairs with single routes, reducing travel time involves

45

coordinating flows on different links of the route. However, when multiple routes exist, the control actions are more globalized and deal with distributing demand appropriately on various routes.

- Reduce link travel time:

    Links are directional roadways that connect points such as intersections, and sources or sinks that discharge and receive vehicles. The travel time on a link is defined as the time taken by a vehicle to travel from the beginning to the end of the link. As the definition suggests, the control actions taken for this objective are localized. This objective attains importance when the link is subjected to abnormal traffic conditions.

- Reduce the time spent by vehicles in the queues:

    Queues may be formed due to recurrent[1] or non-recurrent[2] congestion resulting from the demand exceeding the effective capacity of the roadway. The control system should be designed not just to prevent the growth of a queue but also to clear the queue in the least possible time.

Improve Capacity Utilization

The traffic control system should aim to maximize the utilization of network capacity while ensuring a safe and comfortable driving. This is achieved in terms of two objectives.

- Increase throughput:

    Throughput is defined as the number of vehicles passing through a given section of the roadway in a unit time. The throughput will be maximized when the network is utilized to its fullest capacity. Different ways of increasing throughput include balancing demand across routes via route diversion techniques, and balancing density across links via ramp and mainline metering.

---

[1]Congestion that repeats itself at specific times of the day due to heavy traffic conditions is termed 'Recurrent congestion'. In this case, the effective capacity of the roadway remains the same as the original capacity.
[2]When incidents occur, or weather conditions are bad, the effective capacity is less than the original capacity. The congestion occurring under these conditions is termed 'Non-recurrent congestion'.

- Reduce the number of lane changes:

    Lane changes are of two types - discretionary and mandatory. Mandatory lane changes take place in order to ensure that a vehicle stays on its route to the destination or to bypass closed lanes. Discretionary lane changes are made mainly with the motivation of increasing the speed. This type of lane changes may actually increase throughput if a substantial increase in speed is realized. However, the weaving operations involved with a lane change create a disturbance in traffic and may force the following vehicles to reduce their speeds. Lane changes that do not have significant positive consequences are unnecessary and deteriorate the traffic conditions. Hence, it is the objective of the control system to ensure that the traffic is moving in a way that reduces unnecessary lane changes. For the safety point of view, all types of lane changes should be minimized (this is discussed in more detail under the "Improve Safety Conditions" goal).

Reduce Carbon Monoxide Emissions and Exposure

This goal captures the effect of driving on the environment and the drivers. Emissions in both open and closed sections (like tunnels) have an equally deteriorating effect on the environment. However, emissions in closed sections pose higher risks and health hazards to the drivers than emissions in open sections. Carbon Monoxide (CO), which drives out oxygen from the bloodstream, poses the most serious emissions problem. An extended period of exposure to high CO levels can have serious effects on the behavior of the drivers and their safety. Hence, measures should be taken to reduce the exposure of drivers to harmful levels of CO. There are three ways of reducing the level of harmful CO gases.

- Reduce link density:

    The amount of CO generated is directly proportional to the number of vehicles using the roadway. Hence, by restricting the access of vehicles to critical sections (with regards to CO levels) of the network (and thus, reducing the number of vehicles on a link), we can successfully control the CO buildup. This also helps in reducing the number of drivers that could potentially be affected by the CO levels. Of course, this may conflict with other goals such as "Improve Level of Service".

47

- Reduce the variance of travel speed over time:

  CO emission studies reveal that the acceleration of vehicles results in higher emissions than maintaining a constant speed. Hence, under circumstances of high CO concentration levels, the objective of the control system should also be to ensure that the vehicles in the critical section maintain a uniform speed.

- Increase ventilation fan speed:

  In the covered sections of the network, the CO buildup occurs rapidly. While access and flow control is effective in reducing the emissions subsequently, the existing CO buildup should also be taken care of to minimize the effect on the drivers. Ventilation fans are used to pump out the air with high CO concentration and replace it with fresh air. Under severe conditions, the fan speed could be increased to pump out the polluted air at a higher rate.

Improve Safety Conditions

While goals like improving the level of service and utilizing the network capacity efficiently are necessary for a control system, a safe movement of traffic should also be ensured. Unsafe conditions result in traffic incidents, resulting in deaths and injuries. Accidents also reduce the traffic flow either directly via lane closures or indirectly via "gawkers block" - motorists slowing down to look at the incident. This may, in turn, result in queuing and delays. Hence, the achievement of any goal should not be at the expense of safety. Safety can be achieved by two objectives.

- Reduce the number of lane changes:

  Since a lane change involves acceleration, deceleration, and lateral movement, the disturbance caused is a potential source of accidents. Hence, from a safety perspective, the number of lane changes should be minimized. However, from the point of view of capacity utilization, lane changes made properly could potentially increase the throughput (the objective discussed under the "Improve Capacity Utilization" goal). This is another example where objectives can be conflicting.

- Reduce the speed differential across vehicles on the same link:

  It is improbable to have a large difference in speeds among vehicles on the same lane. However, it has been observed in practice that the differences in speeds among different lanes is quite substantial. This provides an incentive to change

lanes and make travel more hazardous. Hence, the control system should minimize the speed differentials across vehicles on the same link. This not only improves safety but also makes traffic flow more stable.

Maintain Control Stability

The various elements of the traffic network interact with each other. This can render the control unstable. Also, the dynamic nature of advanced traffic control systems results in an unstable oscillatory behavior between the control system and traffic conditions. This is not advisable for a control system. The following objective attempts to alleviate such problems.

• Reduce frequency of signal state changes:

> Frequently fluctuating device states is indicative of unstable control conditions. When this is indeed the case, the frequency of signal state changes should be minimized to improve the stability of the control system.

Control Special Vehicle Access

Special vehicles are grouped under two categories: emergency vehicles such as police patrol, ambulances, and tow trucks; and, trespassing (prohibited) vehicles such as overheight vehicles and vehicles carrying hazardous materials. The emergency vehicles are given priority on some occasions and should be given easy access to the relevant locations in the network during this period. On the other hand, the access of certain locations in the network to trespassing vehicles should be curtailed. The two objectives under this goal deal with these controls.

• Reduce travel time of emergency vehicles from origin to destination:

> Vehicles such as police patrol, ambulances, and tow trucks constitute emergency vehicles. In case of occurrence of an incident, these vehicles should be given easy access to the incident location so that the incident is cleared immediately. Also, emergency care may be needed for people affected by the incident. In most cases, the origin of the emergency vehicle is the nearest emergency station to the incident, and its destination is the incident location. However, once an incident is cleared, the emergency vehicle might be immediately needed in another incident

location. In this case, both the origin and destination of the emergency vehicle will be incident locations.

- Restrict prohibited vehicles from specific locations:

    Vehicles such as overheight vehicles and those carrying hazardous material are prohibited from specific locations in the network. For example, an overheight vehicle approaching a low bridge or tunnel section. When a vehicle of this type approaches a restricted access section, the control system should automatically give a warning and finally ensure that violations do not occur.

3.3.2 Interaction Between Objectives

Each objective is designed to achieve a specific goal. Although it is desirable that the goals are independent of each other, some amount of dependence is inevitable. It is certain to have some overlap in terms of synergy and conflict between goals. For example, by maximizing the utilization of capacity, the level of service is also improved. On the other hand, controlling CO levels in a tunnel or allowing easy access to emergency vehicles may reduce level of service. Similarly, the objectives cannot be designed to be completely independent. Every objective interacts with every other objective, enhancing or hindering its achievement to some extent.

To indicate the interaction between all objectives, a table is constructed. Let N denote the total number of objectives. If the objective representing each column is denoted by {x : x = 1 to N} and the objective representing each row by {y : y = 1 to N}, then the entry in the cell [y, x] indicates the nature of interaction between the objectives y and x. In specific, it signifies the impact of y on x: whether the achievement of objective y strongly or weakly enhances (positive interaction) or hinders (negative interaction) the achievement of objective x. The directionality of interaction should be kept in mind while determining the entries. The entry for cell [y, x] may or may not be the same as [x, y].

The entries in the table reflect only the *a priori* expectation of the type of interaction between objectives. As the evaluation proceeds through the inner loop in Figure 2-3 involving control system configurations, simulation and performance measures (shown in Figure 3-1), the entries will be updated. This will ultimately result in design refinements as represented by the outer loop of Figure 2-3.

50
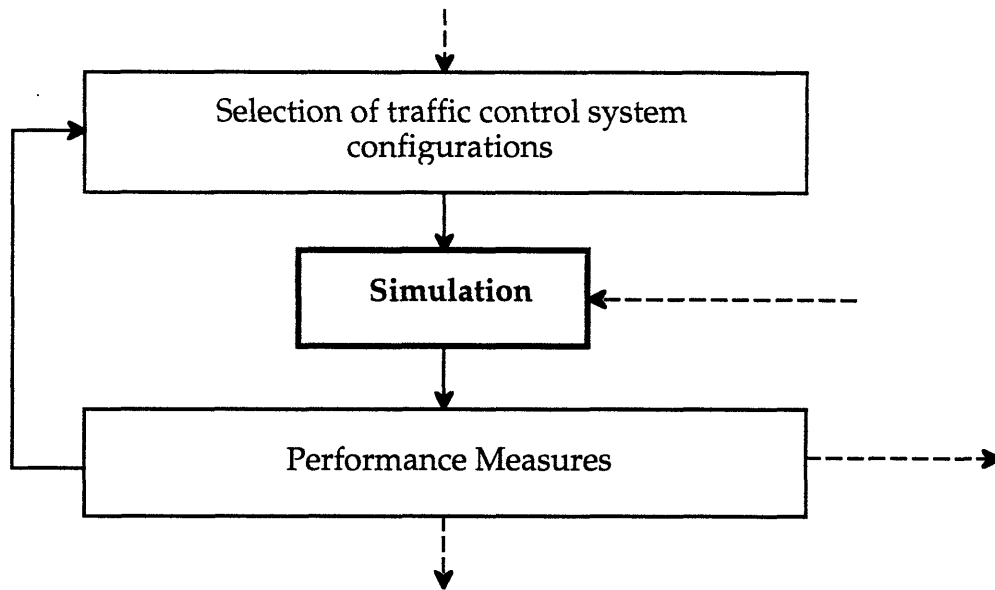
Figure 3-1 The interaction table entries and configuration testing loop (inner loop)

The interaction between objectives has two dimensions:

- the amount of interaction
- the type of interaction.

In some cases, the interaction is strong, while in some others it is weak. On the other hand, the interaction can be either positive (the objectives aid each other) or negative (the objectives conflict with each other). The interaction between two objectives y and x, [y, x] is classified under one of the following 3 categories:

- S+ : Indicates a strong positive interaction i.e., achievement of objective y strongly enhances the achievement of objective x.

- S- : Indicates a strong negative interaction i.e., achievement of objective y strongly hinders the achievement of objective x.

- W : Indicates a weak interaction i.e., achievement of objective y weakly enhances or hinders the achievement of objective x. Since the interaction is only weak, we are not concerned whether it is positive or negative.

Table 3-1 shows the interaction between the various objectives. The following abbreviations are used for the objectives. The first acronym represents the goal and the acronym in parenthesis represents the objective.

51

LOS (TTL)	Improve level of service: Reduce link travel time
LOS (TTOD)	Improve level of service: Reduce travel time on O-D pair
LOS (Q)	Improve level of service: Reduce time spent in queues
CU(TP)	Improve capacity utilization: Increase throughput
CU(LC)	Improve capacity utilization: Reduce number of lane changes
CO(LD)	Reduce CO emissions and exposure: Reduce link density
CO(SV)	Reduce CO emissions and exposure: Reduce speed variance of vehicles over their journey
CO(VFS)	Reduce CO emissions and exposure: Increase ventilation fan speed
SC(LC)	Improve safety conditions: Reduce number of lane changes
SC(SD)	Improve safety conditions: Reduce speed difference between vehicles
CS(SS)	Maintain control stability: Reduce frequency of signal state changes
SV(TTE)	Control special vehicle access: Reduce travel time of emergency vehicles from origin to destination
SV(PV)	Control special vehicle access: Restrict prohibited vehicles from specific locations

The two objectives 'Reduce travel time on OD pair' and 'Reduce link travel time' are combined into one objective 'Reduce travel time' in the table because though the control logic for achieving the two objectives is different, the type of interaction of each of these objectives is the same with every other objective.

The objectives CO(VFS), SV(TTE), SV(PV), and CS(SS) are not within the scope of this thesis and will not be considered further.

Table 3-2 provides an explanation of the interaction between the objective pairs shown in Table 3-1. CU(LC) and SC(LC) represent the same objective and hence, the explanations of interactions of only CU(LC) with other objectives is given in Table 3-2.

Table 3-1 Interaction between objectives

	LOS(TT)	LOS(Q)	CU(TP)	CU(LC)	CO(LD)	CO(SV)	SC(LC)	SC (SD)	TFS(SD)
LOS(TT)	--	S+	S+	W	S-	W	W	W	W
LOS (Q)	S+	--	S+	W	S-	W	W	W	W
CU (TP)	S+	S+	--	W	S-	W	W	W	W
CU (LC)	W	W	W	--	W	S+	--	S+	S+
CO (LD)	S-	S-	S-	W	--	W	W	W	W
CO (SV)	W	W	W	S+	W	--	S+	W	W
SC (LC)	W	W	W	--	W	S+	--	S+	S+
SC (SD)	W	W	W	S+	W	W	S+	--	--
TFS (SD)	W	W	W	S+	W	W	S+	--	--

53

Table 3-2 Explanation of interaction table cells

Objective Pair	Interaction	Explanation
LOS(TT), LOS(Q)	S+	Steps taken to reduce travel time are essentially to ensure that the demand on a section does not exceed the capacity. This automatically results in lesser formation of queues and hence lesser times spent in queues.
LOS(TT), CU(TP)	S+	Successfully reducing travel time allows more vehicles to travel on a route in a given time, which increases throughput.
LOS(TT), CU(LC)	W	Reduction in travel time though lowers the incentive to change lanes does not necessarily reduce lane changes.
LOS(TT), CO(LD)	S-	Lesser time to travel results in higher number of vehicles on the road, which increases link density.
LOS(TT), CO(SV)	W	A reduction in travel time implies a faster travel on the average. This does not necessarily translate into changes in speed.
LOS(TT), SC(SD)	W	Reduction in travel time implies a faster travel on the average. It may be possible that the speeds on different lanes are different. Hence, a higher speed does not necessarily translate to a reduction in speed differential.
LOS(Q), LOS(TT)	S+	A reduction in time spent in the queue implies shorter delays and reaching the destination in a lesser time.
LOS(Q), CU(TP)	S+	A reduction in time spent in the queue implies shorter delays and hence a larger number of vehicles passing through a given section.
LOS(Q), CU(LC)	W	The interaction is very indistinct and definitely not strong. Therefore, a weak relationship is assumed.
LOS(Q), CO(LD)	S-	Reducing time spent in a queue lets more vehicles to pass through the bottleneck section, which increases the number of vehicles on the downstream link.

LOS(Q), CO(SV)	W	The interaction is very indistinct and definitely not strong. Therefore, a weak relationship is assumed.
LOS(Q), SC(SD)	W	The interaction is very indistinct and definitely not strong. Therefore, a weak relationship is assumed.
CU(TP), LOS(TT)	S+	Increasing throughput allows more vehicles to pass through, which reduces the system travel time, on the average.
CU(TP), LOS(Q)	S+	An increase in throughput implies lesser number of vehicles getting stuck/remaining in queues and hence lesser times spent in queues.
CU(TP), CU(LC)	W	An increase in throughput results in higher link density and a more or less uniform speed. This reduces the incentive to change lanes and **may** result in reduction in number of lane changes.
CU(TP), CO(LD)	S-	An increase in throughput automatically increases the link density.
CU(TP), CO(SV)	W	Increased throughput **may** result in a larger number of vehicles on the road. In such a case, a vehicle will not have a chance to accelerate and decelerate freely and hence will maintain its uniform speed
CU(TP), SC(SD)	W	**If** increased throughput results in larger number of vehicles on the road, it **may** not permit the vehicles to travel at different speeds.
CU(LC), LOS(TT)	W	The interaction is very indistinct and definitely not strong. Therefore, a weak relationship is assumed.
CU(LC), LOS(Q)	W	Lane changes made improperly **may** create queues and cause disturbance that results in longer times spent in queues.
CU(LC), CU(TP)	W	Reduction in the number of lane changes reduces the disturbance caused which may increase throughput. On the other hand, it also reduces the possibility of speed increases due to lane changes and may reduce throughput. On the whole, the interaction is assumed to be weak.

CU(LC), CO(LD)	W	Reducing the number of lane changes **may** result in a smoother flow and hence may increase link density.
CU(LC), CO(SV)	S+	Reducing lane changes ensures that the vehicles travel in the same lane. This means that the vehicle cannot increase or decrease its speed randomly.
CU(LC), SC(SD)	S+	Lane changes are reduced when all vehicles travel at a more or less uniform speed. This reduces the speed differential between vehicles.
CO(LD), LOS(TT)	S-	Reducing link density requires impeding vehicles before they enter the link. This results in an increased system travel time, on the average.
CO(LD), LOS(Q)	S-	Reducing link density requires impeding vehicles before they enter the link. This not only increases the number of vehicles getting stuck in the queue but also the time spent by each vehicle in the queue.
CO(LD), CU(TP)	S-	Reducing link density, automatically reduces the number of vehicles passing through a section of the roadway and decreases throughput.
CO(LD), CU(LC)	W	Lesser number of vehicles on the link, may let the vehicles to travel fast and hence, reduce the incentive to change lanes. On the other hand, lesser number of vehicles implies longer gaps which **may** increase the number of lane changes.
CO(LD), CO(SV)	W	Since there will be lesser number of vehicles on the link, the vehicle **may** not accelerate or decelerate frequently. On the other hand, more space **may** result in more maneuvers and speed changes.
CO(LD), SC(SD)	W	Since there will be lesser number of vehicles on the link, they will be traveling at more or less uniform speeds. On the other hand, more space **may** result in vastly differing speeds of vehicles.
CO(SV), LOS(TT)	W	Reduction in the speed variance of a vehicle **may** increase or decrease its travel time or keep its travel time the same.

CO(SV), LOS(Q)	W	The interaction is very indistinct and definitely not strong. Therefore, a weak relationship is assumed.
CO(SV), CU(TP)	W	Same as for CO(SV), LOS(TT)
CO(SV), CU(LC)	S+	Keeping the speed variance of a vehicle at a minimum ensures that the vehicle travels at a more or less uniform speed throughout its journey. If it has to travel at a uniform speed, it does not need to make any lane changes.
CO(SV), CO(LD)	W	A lesser speed variance for all vehicles implies a more stable flow, which **may** result in an increase the link density.
CO(SV), SC(SD)	W	A lesser speed variance implies a more stable flow, which **may** result (not necessarily) in a reduction in the speed difference between vehicles.
SC(SD), LOS(TT)	W	Reducing the speed difference between vehicles results in a uniform flow of vehicles. The resulting average speed may be lower or higher than the previous speed. This may increase or decrease the travel time respectively.
SC(SD), LOS(Q)	W	The interaction is very indistinct and definitely not strong. Therefore, a weak relationship is assumed.
SC(SD), CU(TP)	W	Reducing speed differential between vehicles ensures a uniform speed of the traveling vehicles. This **may** possibly help in increasing the throughput.
SC(SD), CU(LC)	S+	If the speed differential between vehicles drops, then there will be no incentive to change lanes. Hence, the number of lane changes will reduce drastically.
SC(SD), CO(LD)	W	Reducing the speed difference between vehicles results in a uniform flow of vehicles. A more uniform flow **may** possibly result in a higher link density.

SC(SD), CO(SV)	W	Reducing speed differential between vehicles ensures a uniform speed of the traveling vehicles. This will reduce the chances of accelerating/decelerating of individual vehicles. However, if the vehicles are moving as platoons, then, compression waves may occur resulting in an increase in the speed variance of a vehicle.

3.3.3 Control Functions and their Use in Achieving the Objectives

A control function represents a control tool that can be used in various ways to achieve various objectives. This subsection provides a description of the control functions identified to achieve the objectives described in subsection 3.3.1. Four main control functions are identified as part of the traffic control system:

- Route Diversion
- Access Control
- Mainline Control
- Arterial Control.

In what follows is a brief description of each of these control functions.

Route Diversion

A driver usually starts a journey with an initial route planned to be followed. However, while in transit, due to unforeseen disturbances on the chosen route, the driver may be guided or diverted to another route. This is termed "route diversion". The criteria used for the diversion logic depends upon the objective that the control function is aimed to achieve.

Access Control

Under certain traffic conditions, the entry of vehicles to the mainline facility needs to be controlled. For example, when the traffic density on the mainline is high, the access to the mainline section may have to be controlled such that the negative impact of the entering vehicles on traffic flow is reduced. This function serves the purpose of controlling the access to the mainline (tunnel and freeway) facilities and helps in the

elimination, or at least the reduction of congestion. The most widely used form of access control is entrance ramp control.

Mainline Control

While "access control" involves controlling the access to the mainline, "mainline control" deals with controlling the traffic that is already on the mainline. Mainline control then, is concerned with the regulation of traffic on the main freeway and tunnel lanes in the network and may take the form of congestion advisories, route diversion, or lane use control.

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 traffic control. Arterial control then, consists of controlling traffic on frontage roads, parallel arterial streets that can be used as alternate routes, cross streets that link freeway ramps with alternate routes, and the local urban network.

Use of control functions in achieving the objectives

In subsection 3.3.1, the objectives to be achieved to reach each goal have been identified. The control functions identified in this subsection are utilized in achieving these objectives. A control function design specifies a set of algorithms designed for setting the states of certain control devices in a manner aimed at achieving a particular control objective. Though there are four types of traffic controls, not all of these are useful in achieving every objective. The subset of traffic controls that can potentially be used for an objective need to be identified. Table 3-3 shows the control functions that can be used for achieving each objective.

Designing the set of controls to be used for a control function independently to achieve an objective is simple and straight forward. When more than one control function is used - which is the case in integrated systems - they have to be designed to aid each other in order to maximize the benefits and minimize the side effects. Moreover, the control function design will differ based on the objective and the criteria used for the

Table **3-3** Control functions used for objectives

Objective	Route Diversion	Access Control	Mainline Control	Arterial Control
Improve level of service				
Reduce travel time on OD pair	√			√
Reduce link travel time	√	√	√	√
Reduce time spent in queue	√	√	√	√
Improve capacity utilization				
Increase throughput	√	√	√	√
Reduce no. of lane changes		√	√	
Reduce Carbon Monoxide Emissions and Exposure				
Reduce link density	√	√	√	√
Reduce variance of speed over time		√	√	
Increase ventilation fan speed				
Improve safety conditions				
Reduce no. of lane changes		√	√	
Reduce speed diff. across vehicles		√	√	
Maintain control stability				
Reduce no. of signal state changes				
Control special vehicle access				
Reduce tt of emergency vehicles	√	√	√	√
Restrict prohibited vehicles	√	√	√	

design. For example, route diversion used for reducing travel time on a link is based on travel times on different routes while the same function used for reducing density on a high CO concentration link is based on the maximum allowable CO exposure.

3.3.4 Configurations of Control Functions

In this subsection, a method is suggested to obtain alternative control system designs for evaluation. A traffic control configuration is a combination of control functions linked to specific objectives. As a reminder, a control function linked to a specific objective is referred to as a configuration component. Various configurations are

simulated in different iterations of the inner loop (Figure 3-1). These configurations are evaluated in order to assist in refining the traffic control system design regarding both the control function algorithms and the communication between the control functions.

As observed before, most of the control functions may potentially be used for each of the objectives. When used independently to achieve an objective, the benefits (in terms of how close we get to the achievement of our objective) from one control function may be higher or lower compared to that from another. Let us define the best control function for an objective to be one that offers the maximum benefits. The addition of more control functions to aid in the achievement of the objective may or may not increase the benefits, and in some instances, may decrease the benefits. Hence, once the best control function is chosen, the incremental benefits from the addition of one, two or all of the remaining control functions that could be used for the objective should be examined.

The other dimension involved in the evaluation process is the combination of objectives. As observed in subsection 3.3.2, there is some interaction between every pair of objectives. Hence, a control design which is found to be the best for an objective may not be so beneficial when combined with control designs for other objectives. Some other control design which offers lesser benefits when used independently for that objective may give better results when used in conjunction with control designs of other objectives. Also, the knowledge of the interaction between the objectives is essential. Let us recapitulate the importance of interaction between objectives in the control system design.

- The weaker the interaction between objectives, the more independently the control functions for the objectives can be designed.
- When there is a strong interaction, more attention needs to be paid in the design of control functions, especially with regard to the communications between the functions.
- When the strong interaction is positive, the control functions should be designed so as to complement each other, to maximize benefits.
- When the strong interaction is negative, the control functions should be designed hand-in-hand to minimize the overlap.

These interactions should be taken into account during the design process to both enhance the positive interaction and reduce the negative interaction. In doing so, the priorities given to the interactions is in the following order: S+, W, S-. It is be preferable to have a combination of objectives that have only strong positive interactions. Because, the

61

objectives would strongly enhance each other, thereby maximizing the benefits. The inclusion of objectives with weak interactions also does not pose too many complications. However, since the comprehensive control system needs to represent all the defined objectives, presence of strong negative interactions are inevitable. Therefore, the design itself should aim at reducing as much as possible the effect of S-.

Therefore, it is necessary to examine the different levels of interaction: interaction between the different control functions used for the same objective; and the interaction between the control functions used for different objectives to design a comprehensive control system. These interactions are analyzed by the loop shown in Figure 3-1. Complete enumeration of all control function component combinations results in a large number of configurations. An important part of the design process is an understanding of the advantages and disadvantages of each configuration component from the point of view of the objectives. This understanding will allow us to refine the control system more intelligently. Hence, in the beginning of this iterative process, only the combinations with one or two configuration components are considered. This allows for quantifying the performance contribution of each component and the interaction effect across the components. The knowledge gained from these results of one iteration is used in better choosing the configurations with more than two components in the subsequent iterations.

In the outer loop of Figure 2-3 (shown in Figure 3-2) delas with the choice and design of control functions for objectives. The feedback from the simulations performed in the inner loop are utilized in modifying the algorithm designs of the configuration components and in eliminating unnecessary combinations of configuration components. The design modification should also take into consideration the communication between the various control functions. The evaluation process in the inner loop is repeated using the new configuration components. Since the designs are modified, the interaction effects between various components in the previous iterations of the outer loop are not valid anymore. Hence the configuration testing process in the inner loop will have to be repeated.
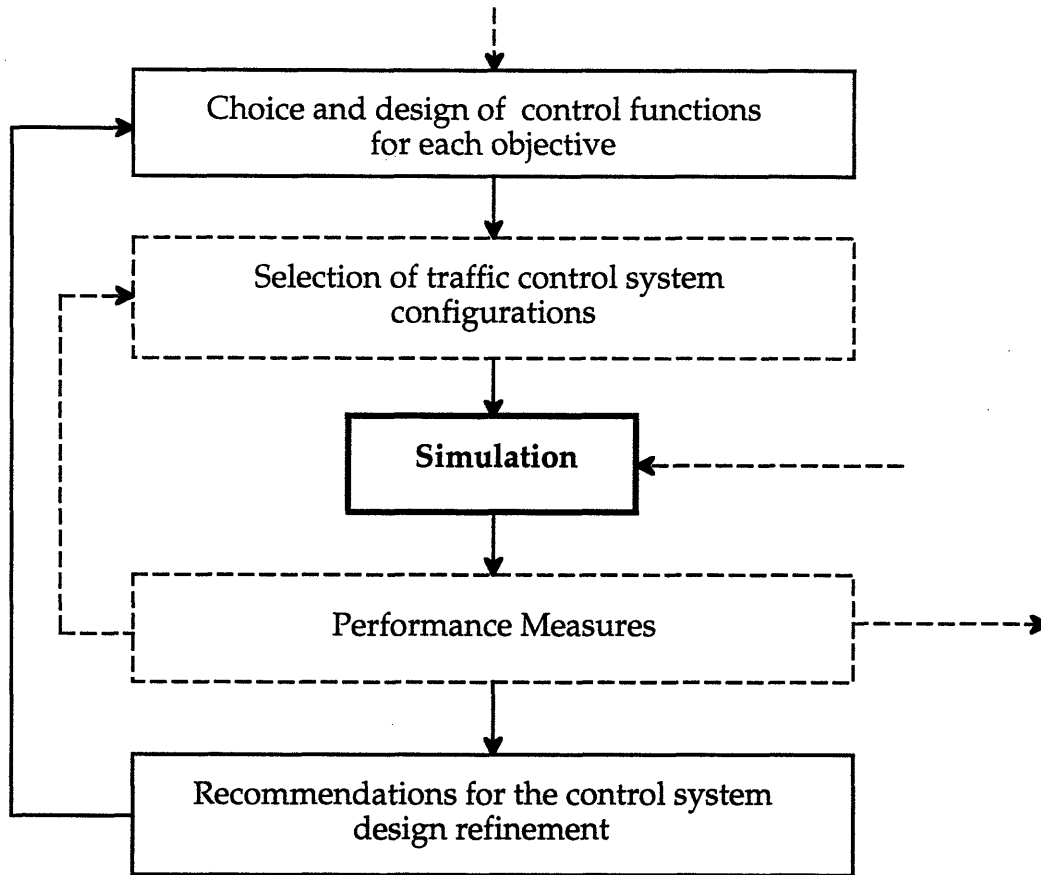
Figure 3-2 The design modification loop (outer loop)

3.4 Scenarios

The scenarios are aimed at testing the robustness of the control system under a wide range of traffic conditions. Scenarios specify the following: traffic volume by origin-destination pair; time, location and severity of events and; the driver behavior parameters. Such scenarios provide a reproducible set of simulation conditions for evaluating the control system design and testing of candidate control configurations. The effectiveness of the response of a control configuration to a scenario is evaluated using the performance measures as described in section 3.5. Each of the control configurations identified for simulation in each of the iterations of the configuration update loop shown in figure 3-1 are tested across a wide range of scenarios. This occurs in the loop involving the scenarios, simulation, and the performance measures in figure 2-3.

This section describes the scenario generation process used in evaluating a traffic control system. In subsection 3.4.1, the groundwork is laid by defining a scenario. The various factors affecting the scenarios are also considered in detail. Subsection 3.4.2 describes a method for the generation of scenarios for simulation.

3.4.1 Definition and Specification

Scenarios are defined by three categories of variables: demand, events, and behavior. Each category consists of a set of variables that can take on a range of values. A scenario is defined by a combination of values across the variables of the three categories. Various scenarios are generated to represent the following three classes of traffic conditions:

- Scenarios that represent normal congestion or free-flow traffic conditions;
- Scenarios that represent moderate traffic congestion resulting from relatively high demand levels or minor events such as a single blocked lane.
- Worst case scenarios that are characterized by severe congestion or oversaturated traffic conditions. These scenarios may result from very high demand levels or major accidents.

In what follows is a description of the categories and the corresponding variables.

Traffic Demand

Traffic demand comprises the first category of variables. These variables determine the amount and composition of traffic that is present on the roadway. Travel demand is specified by time-dependent O-D matrices. The entries in these matrices represent the number of vehicles of each vehicle type that originate within a given interval of time from a particular source in the network with an intended destination at a sink in the network. Different combinations of time dependent O-D demands represent a wide range of possible travel demand patterns.

The vehicle types can broadly be classified under the following classes:
1. Single Occupancy Vehicles
    - High performance passenger cars (New cars)
    - Low performance passenger cars (Old cars)

2. High Occupancy Vehicles

- High performance passenger cars (New cars)
- Low performance passenger cars (Old cars)
- Buses

3. Trucks

- Heavy single-unit trucks
- Trailer trucks.

Vehicles of the same type are assumed to possess common characteristics such as size, maximum speed and desired speed profiles, acceleration and deceleration performance and emission characteristics. The vehicles may also differ in terms of their information availability and communication capabilities.

Traffic demand is specified in terms of Origin-Destination (O-D) tables. The traffic demand is represented as follows:

d^v_{ijk} = demand measured by number of vehicles per hour

where,

v = vehicle type

i = origin of the vehicles ($i = 1$ to N_o, the number of origin nodes)

j = destination of the vehicles ($j = 1$ to N_d, the number of destination nodes)

k = time period ($k = 1$ to K, the number of time intervals in the simulation period).

This represents the elemental entity in the specification of a demand pattern. Figure 3-3 shows a schematic representation of a demand pattern. The matrix d^v_k specifies the number of vehicles of type v going from each of the origins to each of the destinations in the network during the time interval k. Finally, d_k represents an entire demand pattern for time interval k.

65
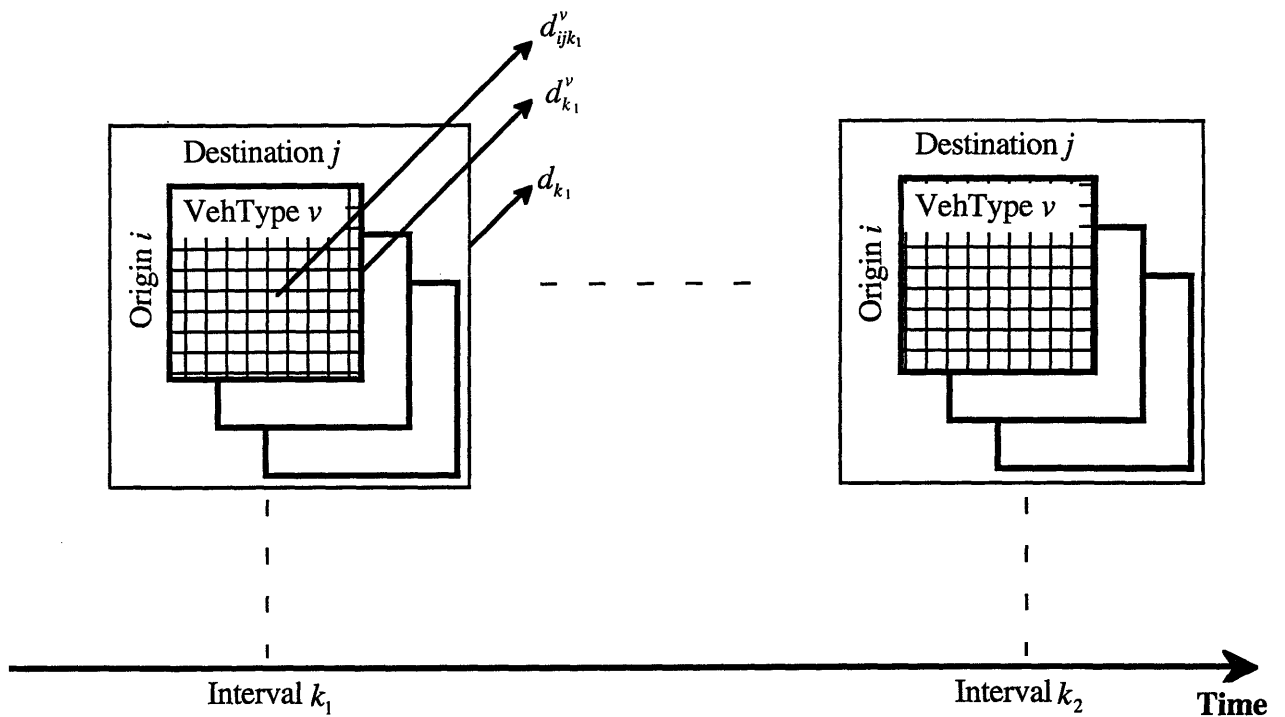
Figure 3-3 Schematic representation of a demand pattern

Events

Events are the second category of variables that make-up a scenario. An event is an occurrence which disrupts traffic flow. Events are not simulated, they are generated exogenously with respect to the simulation. Events include accidents, vehicle breakdowns, hazardous or non-hazardous debris or spillage on the road, fire, weather events, and equipment failures. Potential consequences include lane closure, non-recurring congestion, degradation in surveillance data, and faulty controls. An event can usually be correlated with rapid build-ups of densities in certain lanes and corresponding decreases in volumes and speeds, and a rapid deterioration of air quality in the tunnels. The severity of an event can be judged from the effect it has on traffic flow conditions.

Events are classified under one of the following three categories:
- Accidents: Include vehicle breakdowns, spillages, collisions, and fires
- Device failures: Includes sensor and control device failures
- Weather events: Includes external effects such as rain, snow, and sleet.

The manner in which the various types of events are represented is by setting the values of a set of *event characteristics*. In what follows is a description of the various characteristics and, consequently, an example on how the characteristics can be used in representing a certain type of event.

The Location of the Lane Blockage on the Network

The location of the lane blockage on the network partly determines its impact on traffic flow. For example, blockages close to ramps create spillovers on to local connecting roads.

Number of Lanes Blocked and the Pattern of Blockage

The most common characteristic of a traffic accident is lane blockage. The magnitude of the resulting reduction in capacity depends not only on the number of lanes blocked but also on which lane or lanes are blocked. For example, a double lane blockage is relatively severe compared to a single lane blockage. Similarly, blockages in middle lanes disrupts traffic flow in all lanes more so than blockages in the outer lanes.

The Length of Lane Blockage

The severity of an incident also depends on the blocked 'distance along the lane'. A larger distance causes more disruption to the traffic. For example, events like an oil spill or debris may block a major portion of a lane/lanes while events like vehicle breakdowns may take up very little space. In the former case, the chances that ramps are affected are higher.

The Location of Faulty Surveillance and Control devices

Sensor device failures will result in either lack of or erroneous surveillance data. Moreover, failures of traffic control or information devices may result in improper signal and sign states being conveyed to motorists. The failure of the devices affects the parts of the network influenced by the devices.

Event Duration

The longer the event duration, the larger the effect on traffic flow. Event duration depends upon the severity of the event, the resources available to respond to the event, and the speed of the response (this characteristic relates to both traffic accidents and device failures).

Visibility and Traction Level

Adverse weather conditions can cause reduced visibility and traction. Under such conditions, the motorists are forced to reduce their speeds even when the traffic volume is low. Such conditions affect the reaction time, reaction distance to the signals and consequently requires a larger headway. Other than weather conditions, low visibility can be caused by smoke from a fire and reduced traction may be caused by oil spills and worn out pavement surfaces.

As an example demonstrating the use of event characteristics to represent events, consider the representation of an event involving 'fire'. This event may be simulated as a combination of the following characteristics:
- Number of lanes blocked and pattern of blockage
- Length of lane blockage
- Location of lane blockage
- Any effect on the surveillance and control system, such as location of faulty devices
- Event duration
- Loss of visibility.

An event is represented as follows:

e_{lst}^{c} = value of the characteristic c of the event of interest

where

l = event type (accident, device failure, or weather event)

s = index specifying the location of the event in terms of the link id (each link in the network is assigned a link id number) and the position relative to the beginning of the link.

t = time of occurrence of the event. These are time points at which events occur and not preset time intervals.

68

c = index for characteristic type. The characteristics depend upon the type of event. For example, for accidents, characteristics such as duration and the lanes blocked are specified while for device failures, characteristics such as duration and mode of failure are specified. Mode of failure is 'off' if the device is not functioning and 'erroneous' if the device is faulty.

This represents the elemental entity in the specification of an event pattern. Figure 3-4 shows a schematic representation of an event pattern. A single event of type l is represented by a row vector in the matrix of event l. The vector e_{lst} specifies the characteristics of the event. The number of rows specifies the number of events of type l occurring at time t. Each matrix e_{lt} defines all events of type l occurring at time t. Finally, e_t represents an entire event pattern for time point t. By choosing a set of e's occurring at different times of the simulation period, we specify an event pattern.
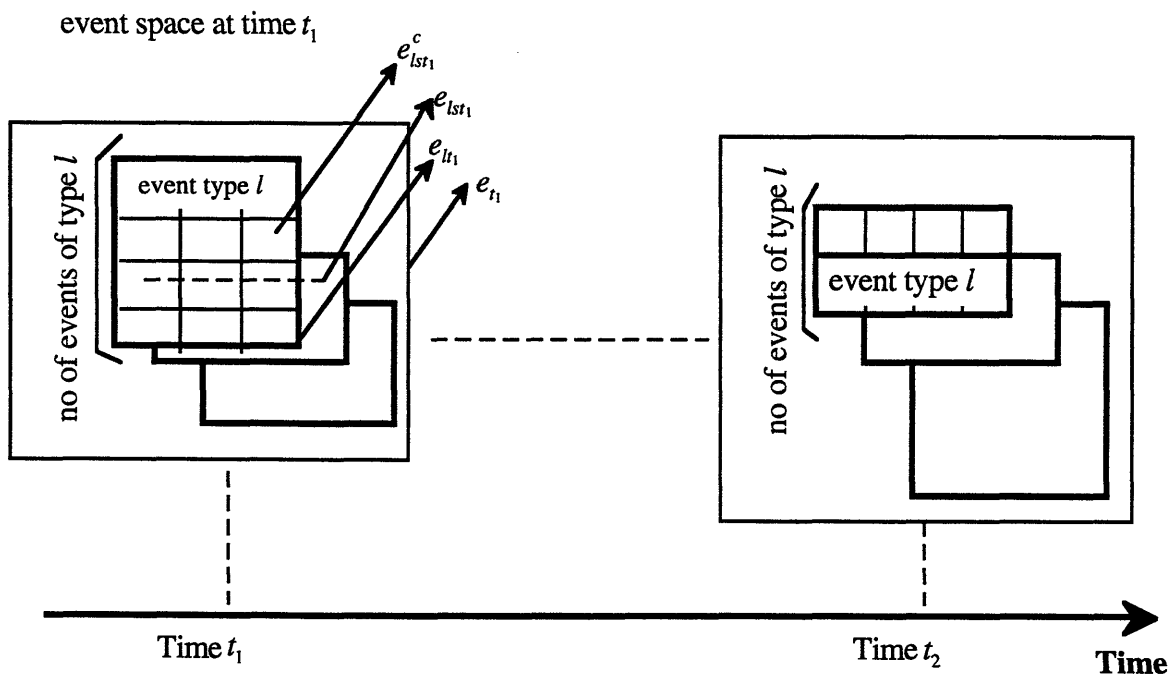
Figure 3-4 Schematic representation of an event pattern

69

Behavioral Parameters

The third and final set of variables that define a scenario are behavioral parameters. These parameters capture the complex behavior of the driver-vehicle pair under different circumstances. Since the behavior of the driver-vehicle pair varies, most parameters are drawn from a probability density function reflecting differences across the population. In such cases, the driver behavior parameters represent the parameters of the probability density function. Since point estimates of all the parameters of interest may not be available at a satisfactory level of accuracy, a reasonable range for the behavioral parameters are captured by the scenarios so that the control system's robustness to varying behaviors can be tested. The following is a description of the behavioral parameters of interest [Yang 93].

<u>Desired Speed Ratio</u>

Speed limit signs are usually not strictly enforced and followed. Hence, the speed limit serves only as a loose upper bound for the driver's desired speed. The desired speed ratio is defined as the ratio of the driver's desired speed to the posted speed limit. Thus, by definition, it is used in mapping the posted speed limit to a driver's desired speed. This is a random number generated from a probability density function.

<u>Headways for the Car-following Relationship</u>

Car-following relationships among vehicles are classified into three categories:
- A vehicle is far from the lead vehicle (free flowing position);
- A vehicle is following the lead vehicle (car-following position); and,
- A vehicle is too close to the lead vehicle (unsafe position).

The speed and acceleration of a vehicle depends on the category it falls under. Each vehicle is assigned a lower bound (H_{Lower}) and upper bound (H_{Upper}) measured in time headway. When the time headway H, from the lead vehicle is less than H_{Lower}, a vehicle is said to be at an unsafe position and emergency deceleration takes place. When H is greater than H_{Upper}, a vehicle is said to be at a free flowing position where the desired speed is the objective. A decision on accelerating, decelerating, or maintaining current speed will depend on whether the current speed is less than, greater than, or equal to the driver's desired speed. When H is greater than H_{Lower} and less than H_{Upper}, a vehicle is

said to be in a car-following position, and the general car-following model is used for calculating the acceleration rate.

Car-Following Acceleration Rate: α, β, γ, and Δt

When the time headway between two vehicles is within the following vehicle's *car-following range* (i.e., $H_{Lower} < H < H_{Upper}$), the interaction with the leading vehicle dominates the following vehicle's decision on accelerating, decelerating or no changing of speed. The acceleration rate for the following vehicle is given by:

$$a_n^{c \cdot f}(t + \Delta t) = \alpha \frac{[v_n(t)]^\beta}{[x_n(t) - x_{n-1}(t) - L_{n-1}]^\gamma} [v_{n-1}(t) - v_n(t)] \tag{3.1}$$

where,

n = following vehicle

$n-1$ = lead vehicle

v_n = speed of the following vehicle

v_{n-1} = speed of the lead vehicle

x_n = position of the following vehicle with respect to the end of the lane

x_{n-1} = position of the lead vehicle with respect to the end of the lane

t = time

L_{n-1} = effective length of the lead vehicle

α, β, γ = parameters

Δt = a stochastic variable describing the response lag for a vehicle.

Lane Changing

The lane changing process involves two different types of behavior:

    - the behavior of a driver performing a lane change, and

    - the behavior of a driver yielding to another driver.

Mandatory and discretionary lane change behaviors are observed from the perspective of the driver performing the lane change. The courtesy yielding process is observed from the perspective of the driver in the target lane (i.e., the driver yielding to another vehicle).

If the current lane on which the vehicle is traveling does not take it to its next link, it is mandatory for the vehicle to change its lane. This is termed as *mandatory lane changing*. The other cases when mandatory lane changing is required is when the road

ahead is blocked due to some event or when the lane use signs are red. The concerned parameters determining the probability of drivers conducting a lane change will change as they approach the point at which the lane change has to be necessarily performed.

The lane changing for the purpose of increasing speed is called *discretionary lane changing*. This is simulated by considering an individual's impatience to a slow leading vehicle. An impatient driver and a slow leader are determined by their current and desired speeds, and the impatience coefficient of each driver. The impatience coefficient is an attribute variable representing the driver's characteristics.

In both types of lane changes, a driver undergoes two stages. The first stage is called the *triggering stage*. During this, a driver *decides* whether or not to lane change. A lane change decision is made if certain conditions are satisfied. These conditions depend upon the type of lane change performed. After the triggering stage, the driver needs to find an acceptable gap size to move into when making a lane change. Once such a gap is found, the lane changing decision is executed.

The triggering stage for mandatory lane changes

The probability that a driver chooses to begin a mandatory lane change is expressed by:

$$p_n = \begin{cases} e^{-\rho x_n}, & x_n \geq \delta \\ 1, & \text{otherwise.} \end{cases} \qquad (3.2)$$

where,

n = vehicle

x_n = distance of vehicle n from the downstream node

ρ, δ = parameters.

ρ is the rate at which the urgency of lane change increases as distance to the target point decreases. Once a vehicle is within δ of the target point (such as exits and positions of red LUSs), the driver will *decide to* change lanes with probability 1. Prior to arriving at this point, the driver will consider changing lanes with probability p_n.

The triggering stage for discretionary lane changes

One of the conditions that need to be satisfied for a driver to perform a discretionary lane change is that the driver's current speed must be much lower than the desired speed. The current speed (v_n) must be less than the desired speed (V_n) by some factor:

$$v_n < \zeta_n V_n \tag{3.3}$$

where,

ζ_n = impatience factor.

The Gap Acceptance Stage

Two buffer time headways influence the execution of a lane change:
- the lead gap buffer time headway, h^{lead}: the minimum lead headway the driver accepts in lane changing, and
- the lag gap buffer time headway, h^{lag}: the minimum lag headway the driver accepts in lane changing.

The buffer time for lead and lag gap of lane changing may differ between individuals and, therefore, should be modeled as a random variable as well. The lead and lag gaps are specified separately for each type of lane change.

Courtesy Yielding

In some situations a driver either decelerates or accelerates using a lower rate than that could be used in order to permit another driver to merge into his or her lane. This is termed as Courtesy Yielding. The probability of a vehicle yielding to another is specified separately for discretionary lane changing and mandatory lane changing. The probability of yielding to a mandatory lane change is higher than that to a discretionary lane change because the yielding driver is more aware of the lane change situation. The probability of courtesy yielding is represented by p.

Merging Headway Buffer

When two or more upstream lanes merge into a downstream lane (for example, at freeway on-ramps or after toll booths), several vehicles may compete for a downstream

73

lane. In such a case, a safe merging check is required. Based on the *merging constraint*, a merging headway buffer between vehicles is obtained.

<u>Startup Delay</u>

When a platoon of vehicles is subjected to a kinematic disturbance, the speed of vehicles will decrease gradually and finally they may come to a complete halt. After the lead vehicle moves, the next vehicle will move after a few seconds of delay. This delay is referred to as the 'startup delay'.

The behavioral parameters are specified by b, a vector containing all the parameters. These sets of behavior parameters assist in making judgments as to their impact on the traffic conditions and relative changes that could be applied to the control system to improve the traffic conditions. If a set of parameters worsens the traffic conditions, the amount of deterioration and the type of changes that could be imposed on the control strategies under these conditions to counteract this effect can be studied.

Scenarios

A scenario is a combination of a demand pattern, an event pattern and a vector of behavioral parameters. A scenario can be mathematically represented as $\{d, e, b\}$ where,

d = the demand pattern: $\{d_1, d_2, \ldots\ldots, d_K\}$

e = the event pattern: $\{e_1, e_2, \ldots\}$

b = behavioral parameter vector

3.4.2 Scenario Generation

Considering a scenario to be a point in a three-dimensional space with each of the three factors forming a dimension, the three-dimensional space can be represented as in Figure 3-5. Any point in this space represents a scenario. It should be kept in mind that the combination of the factors should be reasonable enough to be called a possible scenario. Also, the objective of the scenario generation is to create and choose scenarios that will challenge the control system the most. Because only under such conditions can the robustness of the control system be properly tested. Such scenarios offer the best opportunity to detect deficiencies, if any, in the control system. Hence, the event patterns should be chosen in correlation with traffic demand and the behavioral parameter sets. In

74
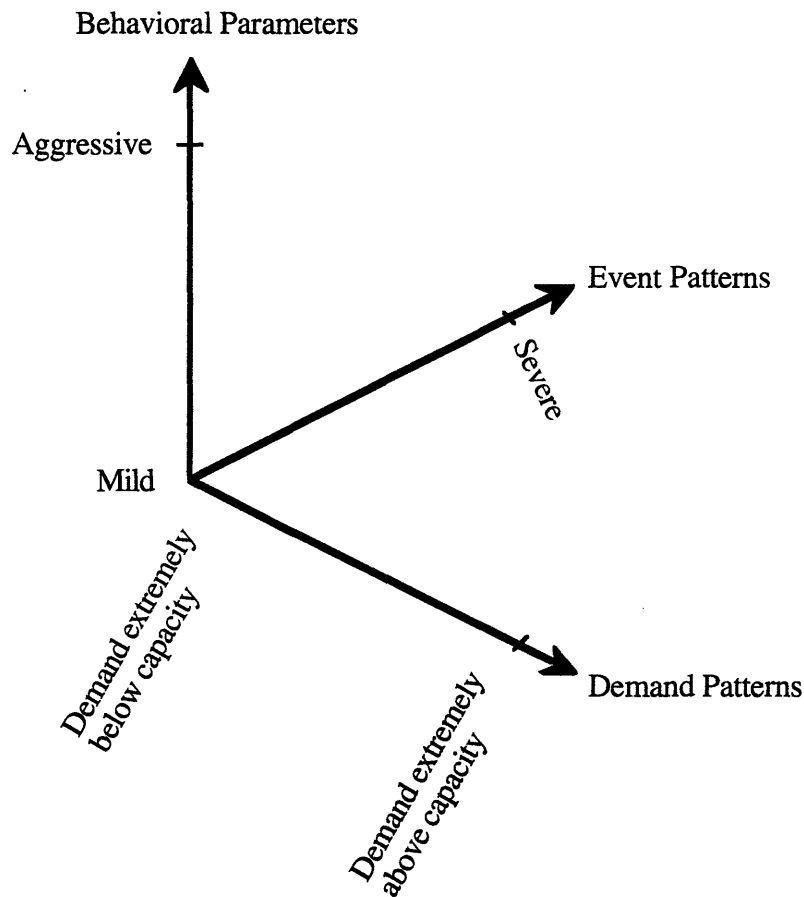
Figure 3-5 Graphical representation of the three factors of a scenario

spite of eliminating any unreasonable scenarios, given the range of variability of the various parameters affecting each factor, we obtain a prohibitive number of scenarios. The number of scenarios may be reduced by following the method of scenario generation described below.

The proposed method follows an iterative approach of generating scenarios as illustrated in Figure 2-3 by the loop involving scenarios, simulation and the performance measures (shown in Figure 3-6). Initially, a scenario is generated and the results from the simulations are utilized in refining the scenario for subsequent simulations.
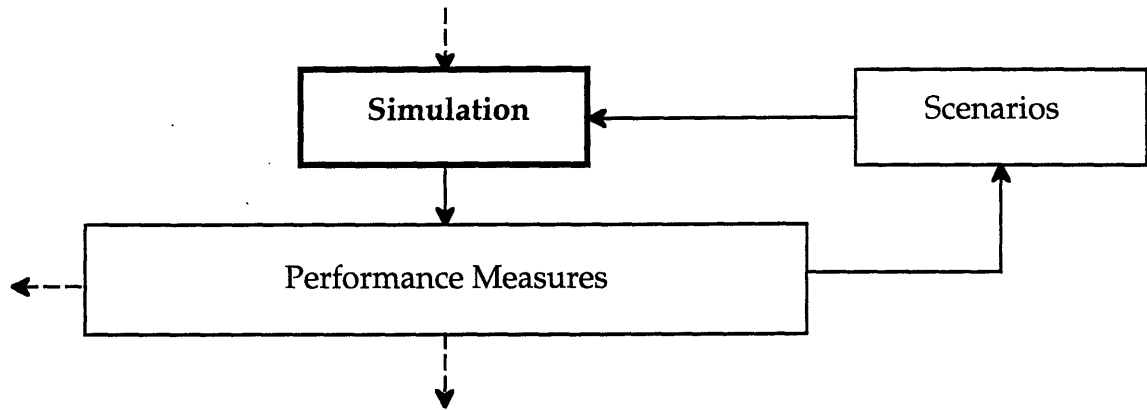
Figure 3-6 The Scenario Generation loop

Initially, a comprehensive set of demand patterns for the simulation need to be generated. Traffic demand on the network exhibits a wide variation, especially depending upon the time of day. The traffic volumes vastly differ between AM and PM periods and peak and off-peak periods. Also, special events alter the daily average volumes drastically. The traffic patterns chosen for simulation should reflect these variations.

A network is designed or modifications made to its existing design to meet the demands of a future year. The volumes on the various O-D pairs for the target year are forecasted as a function of the existing volumes and the effect of the new design. These forecasts are based on a number of assumptions. Hence, they are not always exact and have a margin of variation. Thus, a few demand patterns for different time periods are available from the planning agencies. Extra demand patterns for use in the simulation can be obtained as follows based on special considerations:

- Relaxing or making an assumption more restrictive.
- Higher vehicle volumes or higher mix of heavy vehicles on specific O-D pairs
- Higher vehicle volumes or higher mix of heavy vehicles departing from a specific origin.
- Higher vehicle volumes or higher mix of heavy vehicles reaching a specific destination.

An additional variable that could be manipulated with to obtain a new pattern from the last three considerations is the time period. Specific time periods could be marked during which the demands are varied. Thus, a large number of demand patterns can be obtained for use in the scenarios.

Similarly, all sets of behavioral parameters representing a reasonable spectrum of behavior is generated. Modeling the behavior of the drivers realistically requires a number of parameters as detailed in the subsection 3.4.1. Given the number of parameters and the range of values over which each of the parameters can be varied, a large number of sets are possible. To reduce the number of behavioral parameter sets to a manageable number for the simulation while still capturing the range of interest, a methodological approach is followed in choosing the sets. This approach is described next.

Initially, the sensitivity of each of the performance measures to changes in behavioral parameters is determined. This information can be obtained from a sensitivity analysis. A table is formed that illustrates the effect of the change of each of the behavior parameters on each of the measures, with the format as shown below.

	Measure1	Measure2	Measure3	Measure4	Measure5	
Beh. Par. 1						
Beh. Par. 2						
Beh. Par. 3						
Beh. Par. 4						
Beh. Par. 5						

The effect of the parameter on the performance measure is classified into four categories.

- S+: Means that an increase in the value of the parameter strongly improves the value of the measure i.e., the proportion of improvement in the value of the measure is at least as much as the proportion of increase in the value of the parameter. If a smaller value of the measure is preferred (e.g., cumulative CO exposure), then an S+ means that an increase in the value of the parameter decreases the value of the measure. If a larger value of the measure is preferred, (e.g., average speed), then an S+ means that an increase in the value of the parameter increases the value of the measure.
- S-: Means that an increase in the value of the parameter strongly deteriorates the value of the measure.
- W+: Means that an increase in the value of the parameter weakly improves the value of the measure i.e., the proportion of improvement in the value of the

measure is relatively lesser than the proportion of increase in the value of the parameter.

- W-: Means that an increase in the value of the parameter weakly deteriorates the value of the performance measure.

The entries in the cells of the table will be one of the above four categories. The performance category (reflecting the control goals) under which each of the measures fall is also known. Given all this information, behavioral parameter sets can be chosen by

- focusing upon specific objectives independently (like improving or worsening a specific objective, say, level of service, by making relevant changes to behavioral parameters) or
- focusing upon a set of objectives simultaneously.

That is, each set of parameters could be designed to influence a single objective or multiple objectives. While applying changes to the values of the parameters, it has to be ensured that the behavior of the driver is consistent among all the parameters. That is, the driver cannot exhibit aggressive behavior based on some parameters and mild behavior based on some other parameters. Thus, all the sets of behavioral parameters that are possible and are deemed necessary for simulation are enumerated.

Given the various demand patterns and sets of behavioral parameters, relevant combinations of the two are enumerated. For each of these combinations, an event pattern is necessary to generate a complete scenario. An initial event pattern is generated on the network. Simulations are executed with each of the demand pattern and behavioral parameter combinations and the event pattern.

A complete set of performance measures are obtained at the completion of the simulations. These measures are used to generate a new event pattern to further challenge the control system. This new event pattern will be used in combination with the predetermined demand-behavior combinations. Once a new event pattern is obtained, the simulations are repeated. Iterations continue until the control system is sufficiently challenged.

In subsection 3.3.4, it was mentioned that for each iteration of updating configurations (see Figure 3-1), a set of configurations are chosen a priori. Each configuration is simulated under a wide range of scenarios. For the first simulated configuration, the procedure described in this section generates a set of scenarios. The

78

performance of the configurations can be compared only if the same set of scenarios are used for each one. Therefore, the set of scenarios generated through the first configuration in an iteration are used for testing all configurations of that iteration. However, a new set of scenarios may have to be generated when the control system design is refined via the outer loop of the design-evaluation process (see Figure 3-2). In such a case, the evaluation process within the recommendations for refinements step should take this into account.

3.5 Control System Performance Evaluation

As mentioned before, the inputs to the simulation are the control system designs and the scenarios. The output from the simulation is a set of performance measures. The design of alternative control systems was discussed in section 3.3. Section 3.4 dealt with the definition and generation of scenarios to be used in the simulation. In this section, the performance measures are discussed. The remaining portion of this section is divided into two subsections. In subsection 3.5.1, a set of performance measures is presented and in subsection 3.5.2, different ways of using these performance measures in evaluating alternate control system designs is discussed.

3.5.1 Performance Measures for Evaluation

The performance measures are used to evaluate the control system designs and consequently have to be designed to capture the spatial and temporal impacts of the control system on traffic flow. This subsection provides a description of the performance measures. The definitions of the performance measures with respect to vehicle class, space, and time are presented. Specifically, vehicle class measures are determined by the vehicle type and performance. Space is captured by one of the following six categories: point, segment, link, path, origin-destination (O-D) pair or the whole network. Time is captured by one of the following four categories: instantaneous (snapshot), order of minutes, order of hours, or one day.

This subsection is divided into two parts. The first part describes the control system's goals the various measures are associated with. The second part provides a description of the various measure types, the different classes of measures under each

type, and their significance. The specific measures and their mathematical expressions that could be used for the evaluation are presented in Appendix A.

Performance measures and the control system's goals

The performance measures are grouped under six categories indicative of the goals the control system aims at achieving (see section 2-1). In what follows is a presentation of these categories.

Level of Service

A transportation facility should accommodate a quantity of traffic demand with a quality of service in accordance with the objectives of the system. Indicators of level of service include speed, travel time, traffic interruptions, freedom to maneuver, and queuing.

Capacity Utilization

This goal captures the overall operational efficiency of the system in terms of utilization of capacity, and the amount of travel measured by total vehicle miles and vehicle hours traveled.

Carbon Monoxide Emissions and Exposure

The air pollution problem is related to the characteristics of and use of the vehicles on the road. The interest here is in the extent of emissions and exposure of the drivers and passengers to Carbon Monoxide (CO) in tunnel sections. The operation of the tunnels in a network should adhere to the standards provided by the Environmental Protection Agency (EPA).

Safety Conditions

One of the aims in the design of a traffic control system is to provide a safe movement of traffic. The level of safety provided is measured in terms of speed distribution and lane change frequencies. A uniform speed across all vehicles on a link or

segment reduces the frequency of overtakes and, consequently, creates a safer travel environment. Reducing the frequency of lane changes has a similar effect.

Control Stability

Advanced traffic management systems are adaptive to changing traffic conditions. Inherent to all automated systems is the issue of system stability. It is important that the control system be designed with sufficient stability to accommodate a wide range of traffic conditions. Measures that capture control instabilities include frequency of signal state changes.

Control Special Vehicle Access

The control system should facilitate easy access of affected locations by special vehicles such as police cars and other emergency vehicles so as to restore normal conditions in the least amount of time. This is measured by the travel time of the special vehicles. Similarly, vehicles such as overheight vehicles should be disallowed from entering certain sections of the network. Whether the control system was successful in doing so is gathered by monitoring the movement of these vehicles over their journey.

The performance measures are grouped together by type. For example, all the measures pertaining to travel time are classified under the travel time type. Measures of the same type represent various levels of detail along with temporal and spatial aspects. Table 3-4 shows the association of each type with the control goals. As evident from the table, a specific measure type may be associated with several goals. The reasons behind this will become clear in the subsequent section.

Table 3-4 Categories of performance measures

Measure Type	Performance Category (Goal)					
	Level of Service	Capacity utilization	CO emiss & expos.	Safety Condition	Control Stability	Spec. Veh Access
Carbon Monoxide			√			
Travel Time	√	√				√
Speed	√			√	√	
Lane changing	√			√	√	
Density	√	√				
Queuing	√	√				
Freq of signal state changes					√	
Vehicle Miles Traveled		√				
Vehicle Hours Traveled		√				

Description of the performance measures

The presentation of the performance measures is based on the following structure. For each measure type, the different classes of measures under each type, and the control goals that the measures capture are discussed. Fort each class of measures, the various aggregation schemes used in computing the measures are presented. As discussed above, an aggregation scheme relates to space, time, and vehicle category.

Carbon Monoxide

The measures under this type capture the effect of travel on the environment and drivers. Three classes of measures are obtained from this measure type:
- CO emissions
- CO exposure
- Severity of CO exposure

While the first type captures the effect on the environment, the last two capture the effect on the drivers. Among the different types of gases emitted by vehicles, CO is the most harmful gas. Hence, the focus on CO emissions and exposure.

CO Emissions

Emissions from vehicles affect the air quality. CO emissions are taken as a measure of the overall emissions by all the vehicles in the system over the entire simulation period. The measures derived from this are indicative of the environmental conditions in the network.

CO Exposure

Due to the dispersion of emissions into the open atmosphere and possible ventilation through fans in the closed sections, the exposure of drivers to CO levels may not be the same as level of CO emissions. Hence, CO exposure is specifically considered as a separate measure class. CO exposure is defined as the cumulative CO levels encountered by a vehicle during its journey.

Carbon Monoxide drives out the oxygen in the bloodstream. Small exposure to CO can cause dizziness, fatigue and slow driving reactions and a high concentration over a short exposure time can be a serious health hazard. Moderate concentrations often exist in tunnels and heavy traffic. This class of measures capture the amount of exposure of all vehicles over their journey. It is assumed that the dissipation rate of CO from the blood stream to be zero for the time periods of travel.

Severity of CO exposure

Air quality standards established by the EPA specify the maximum CO levels (in terms of parts per million) that a person can safely experience for different time periods. Figure 3-7 shows the safety limits on CO exposures for different time periods as set by EPA/FHWA [EPA/FHWA 93]. Notice how a smaller concentration of CO levels suffices to reach the safety limits when the time of exposure to CO is long. Similarly, when the CO concentration is high, exposure for a small amount of time is enough to reach the safety limits. The safety limit for cumulative CO exposure is obtained from the figure as the area of any of the boxes bounded by the x-axis, y-axis and the solid line. For example, (a.c = b.d) specifies the safety limits on cumulative CO exposure. This class of measures capture the severity of exposure with respect to the safety limit.
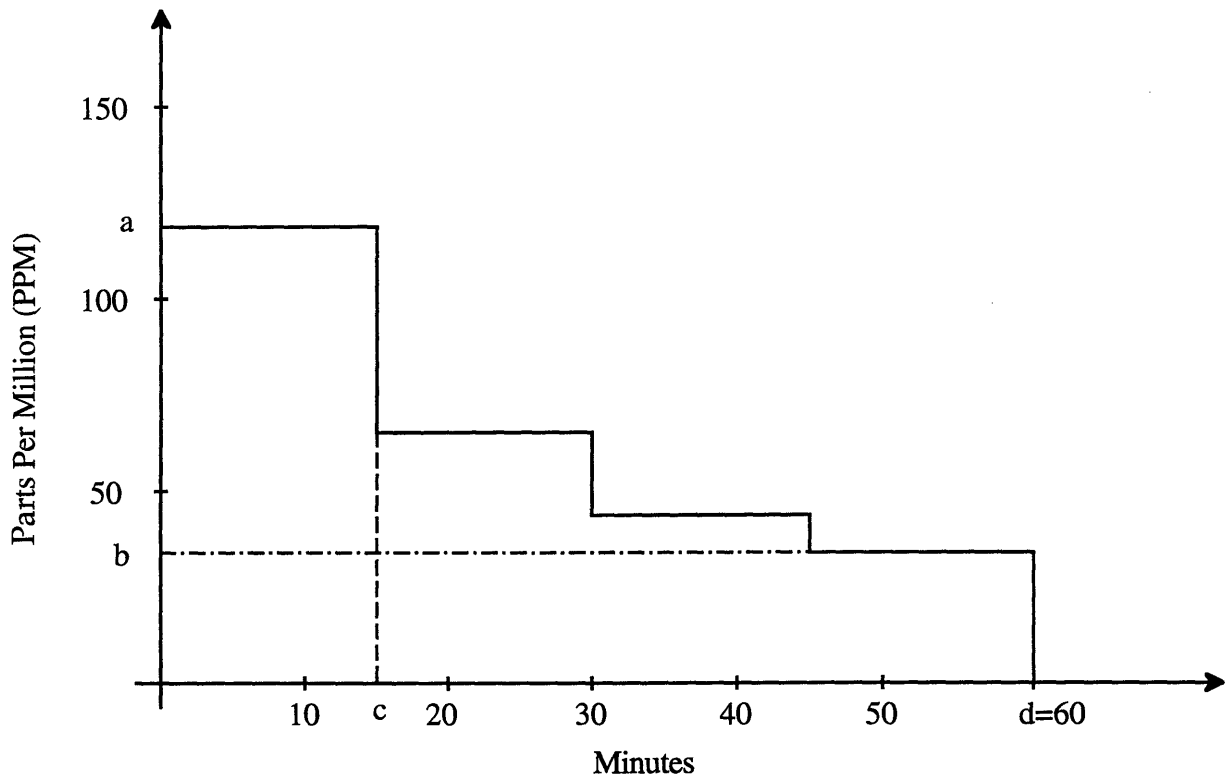
Figure 3-7 EPA/FHWA CO Safety Limits

Travel Time

Three classes of measures are derived from 'Travel Time':

- Link travel time
- Origin-destination travel time
- Travel time imbalances across paths

Each of this is described below in more detail.

Link Travel Time

Link travel time is defined as the time taken by a vehicle to traverse the link (travel from the beginning to the end of the link). The measures derived from this class are indicative of the level of service at local areas on the network. The performance measures are computed for specific links for specific intervals of time without discrimination to the vehicle category.

Origin-Destination Travel Time

Origin-destination travel time is defined as the time spent traveling from an origin to a destination. The actual travel time relative to free-flow travel time captures the level of service provided by the system.

The performance measures are aggregations of the travel times across various variables. These aggregations are based on the data aggregation scheme summarized in Table 3-5. The aggregation is conducted over all the O-D pairs or over all pairs of special interest. Moreover, the aggregation can be over all vehicle categories or a specific category such as High Occupancy Vehicles (HOV). In time, all the vehicles departing either during a specific short period (e.g., the half hour exhibiting the highest demand) or during an extended period are considered.

Table 3-5 Aggregation scheme for travel time

O-D Pairs	Departure Times	Vehicle Category
- all pairs	- extended period	- all types
- pairs of special interest	- peak period	- each category individually

Travel Time Imbalances Across Paths

The travel time imbalance represents the maximum absolute difference in travel times across alternative paths of an O-D pair. The measures derived from this imbalance capture the efficiency of capacity utilization. Initially, for each of the different paths of an O-D pair, the average of the travel time over the vehicles of each category that have departed from O towards D during a 5 minute interval is computed. In cases where a vehicle changes its path enroute, the vehicle observations will be attributed to the realized O-D pair. The travel time imbalance for the O-D pair is defined as the difference in the maximum and minimum of these averages across all the O-D paths for a vehicle category and a time interval. The aggregation scheme is the same as in Table 3-5.

Speed

This measure type contributes to three classes of measures:
- Speeds at point locations

- Traffic speed-speed limit differentials
- Speed during journey

Speeds at Point Locations

The purpose of the set of performance measures under this class is to address safety and control system stability issues. The performance measures relate to the time-mean-speeds at specific locations for short intervals of time (in the order of half a minute). The time-mean-speed is defined as the arithmetic mean of the speeds of the vehicles that pass a given point on the roadway during a given period of time. Such selected points represent locations associated with a relatively high accident risk and include the following:
- downstream of locations where 2 or more roadways merge (such as entrances)
- upstream of locations where a roadway diverges to 2 or more roadways (such as exits)
- upstream of geometric constrictions.

Traffic Speed - Speed Limit Differentials

Measures falling under this class may indicate instabilities in the control system. The performance measures derived from this relate to the difference between the speed limit set for a roadway section and the space-mean-speed of that section measured at a frequency in the order of once every half a minute. The 'Space-mean-speed' is defined as the average speed of vehicles traversing a given length of a road section, weighted by the times spent to traverse the section. The averaging is across all vehicles contained within the section of interest at a specific point in time (a snapshot).

The sections that are considered represent the zones of influence of Variable Speed Limit Signs (VSLS). This zone depends upon the location of the VSLS under consideration, the distance to the next VSLS and the sight distance (the distance at which the signs become visible to the drivers). The zone of influence of a VSLS begins at the point at which it is visible to a driver and ends at the point at which the next VSLS comes into view. As an example, Figure 3-8 shows the zone of influence of VSLS2.
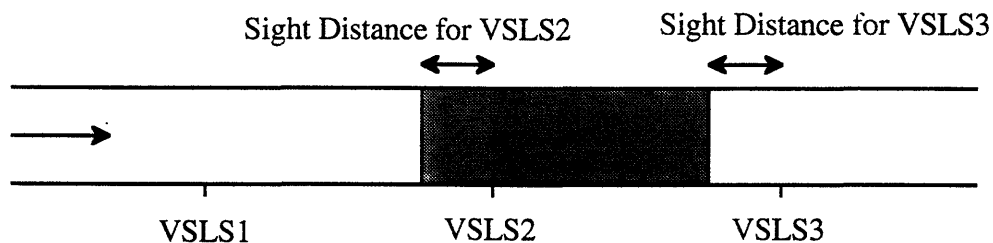
Figure 3-8 Zone of influence of VSLS

<u>Speed During Journey</u>

The speed of a vehicle traced over its journey from origin to destination can be used as a measure of the level of service and safety provided by the system. For example, a higher speed indicates lesser congestion and a higher level of service. However, a high mean speed coupled with a high standard deviation suggests that the vehicle has been frequently accelerating or decelerating rapidly (including coming to stops), which can potentially lead to accidents. Thus, this measure is also indicative of the safety conditions in the network.

Lane changing

Lane changes are of two types: mandatory lane changes and discretionary lane changes. While the former are executed for reasons such as to ensure the drivers are on the appropriate paths to their destinations and to avoid blocked lanes, the latter are made mainly with the intention of increasing speed. Two classes of measures are derived from this measure type:

- Lane changing during journey
- Lane changing at specific locations

<u>Lane Changing during journey</u>

The number of lane changes a vehicle executes during its journey is a level of service and safety indicator. The performance measures are computed independently for each category of vehicle and separately for both mandatory and discretionary lane changes.

87

<u>Lane Changing at Specific Locations</u>

The number of lane changes at specific locations in the network is a measure of safety of the system. The performance measures are computed for specific locations on the network. These locations are critical sections where a number of merging and diverging maneuvers take place. The selected sections include the following.

- downstream of locations where 2 or more roadways merge (such as entrances)
- upstream of locations where a roadway diverges to 2 or more roadways (such as exits)
- upstream of geometric constrictions.

No distinction is made between mandatory and discretionary lane changes.

Density

Density is measured as the number of vehicles per lane-mile within a roadway section at a specific point in time. This measure is indicative of the level of service and efficiency provided by the system. Density is an indirect measure of the space headway and hence, can be used to measure the level of service provided by the system. Also, a high density on one link and a low density on an adjacent link suggests an inefficient utilization of capacity. The performance measures are computed for specific paths at specific points in time.

Queuing

The performance measures derived from this are used to measure the level of service provided by the system and its efficiency. A queue is said to be formed when two or more vehicles line up one behind another and satisfy some criteria. Two different criteria are considered to check if a vehicle is part of a queue:

1. The time and space headways between two adjacent vehicles are less than a defined threshold.
2. The velocities of the vehicles are less than a specific threshold.

If any of these criteria is satisfied for a vehicle, then the vehicle is said to be a part of the queue. The criteria used depends upon the location at which the queue is considered. For example, consider a section upstream of the incident location. For complete lane blockage, since the vehicle cannot pass through until the incident is

cleared, it comes to a halt. The second criterion is used in this context. On the other hand, for a partial lane blockage, the vehicles can still move forward. In this case, the speeds vary frequently and the first criterion is used.

For queuing, an aggregation scheme is defined by the queue type and the vehicle status. Four types of queues are considered: those caused due to partial blockage incidents, complete blockage incidents, due to portal closures, and those formed on ramps. The vehicle status is defined by whether it is exiting on the off-ramp upstream of the incident or portal or moving straight on the freeway. The aggregation scheme is illustrated in Table 3-6.

Table 3-6 Aggregation scheme for queuing

Queue Type	Vehicle Status
- partial blockage incidents - complete blockage incidents - portal closure	- vehicles exiting upstream of incident or portal - all vehicles
- ramp metering	- all vehicles

Two categories of measures are derived from this measure type:
- Lengths and durations of the queues
- Time spent by vehicles in the queues.

Lengths and durations of queues

Queue length is measured as the number of vehicles in the queue. Duration of a queue is defined as the time from the formation of the queue to its complete dissipation. For the measures belonging to this category, the aggregation scheme reduces only to the queue type. No discrimination is made regarding vehicles. The performance measures are computed independently for each queue type.

Time spent by vehicles in the queues

These measures are computed independently for each type of queue and for vehicles of different status.

Frequency of signal state changes

Advanced traffic management systems will employ control systems that adapt to changes in traffic conditions. Consequently, a measure is needed to assess the stability of the control system. The performance measures derived from this relate to the rate at which the changes in the signal states take place. The state changes are aggregated across each type of signal, e.g. across all LUSs, across all VSLSs, etc. The performance measures are computed for each signal type independently.

Vehicle Miles Traveled

Vehicle-miles-traveled is a measure of the distance traveled by all the vehicles in the system over the entire simulation period. The measures derived from this are indicative of the efficiency of the system.

Vehicle Hours Traveled

Vehicle-hours-traveled is a measure of the travel time spent by all the vehicles in the system over the entire simulation period. The measures derived from this are indicative of the efficiency of the system.

3.5.2 Using Performance Measures to Evaluate Control Systems

Section 3.4 addressed the issue of using the performance measures for the generation of scenarios. In this subsection, the focus is on the usage of performance measures in the choice of configuration sets for simulation and to evaluate alternate control system designs.

Subsection 3.5.1 enumerated the various performance measures that are obtained from the simulation of each scenario and control system. The whole set of measures is used for evaluation purposes. As it has been mentioned before, the measures are grouped under categories signifying the various control goals. And the decision regarding the control system design is made based on the performance of alternate designs with respect to the goals.

When there is a single objective, the objective measure obtained from the simulation for one design can be compared to that from another and decisions can be made directly regarding the best design. However, when there are multiple objectives, the decision process becomes more complicated. If a design performs better than another with respect to all objectives, then the problem is simplified, because it is the obvious choice for the best one. However, this is hardly ever the case with alternatives. In most cases, while a design performs better than another under some objectives, it is worse-off with respect to other objectives. Hence, the problem of performing an evaluation of alternate designs when we have multiple objectives is complex and needs a comprehensive analytical approach. The approach should be capable of organizing and comprehending the large amount of data available and be able to identify clearly tradeoffs between conflicting objectives. It should also be able to compare designs systematically and present useful and understandable information on which to base the decisions.

The presence of multiple objectives to base our decisions on result in a multiobjective decision making problem. A description of a few methods that could be used for decision making under these conditions is given in Appendix B. One such method, called *concordance analysis* is described in more detail in relation to the application at hand. It uses pairwise comparisons across objectives to assess the performance of alternate control designs. The rest of this subsection outlines the approach for obtaining a satisfactory design via an iterative evaluation-refinement process. The part of Figure 2-3 involving the configuration updates and control system design refinements is shown in Figure 3-9.

```
        ┊
        ▼
┌─────────────────────────────────────┐
│   Choice and design of control       │
│   functions for each objective       │
└─────────────────────────────────────┘
        │
        ▼
┌─────────────────────────────────────┐
│   Selection of traffic control       │
│   system configurations              │
└─────────────────────────────────────┘
        │
        ▼
┌─────────────────────────────────────┐
│          Simulation          ◄───────────────
└─────────────────────────────────────┘
        │
        ▼
┌─────────────────────────────────────┐
│      Performance Measures            │ ------->
└─────────────────────────────────────┘
        │
        ▼
┌─────────────────────────────────────┐
│  Recommendations for the control     │
│  system design refinement            │
└─────────────────────────────────────┘
```
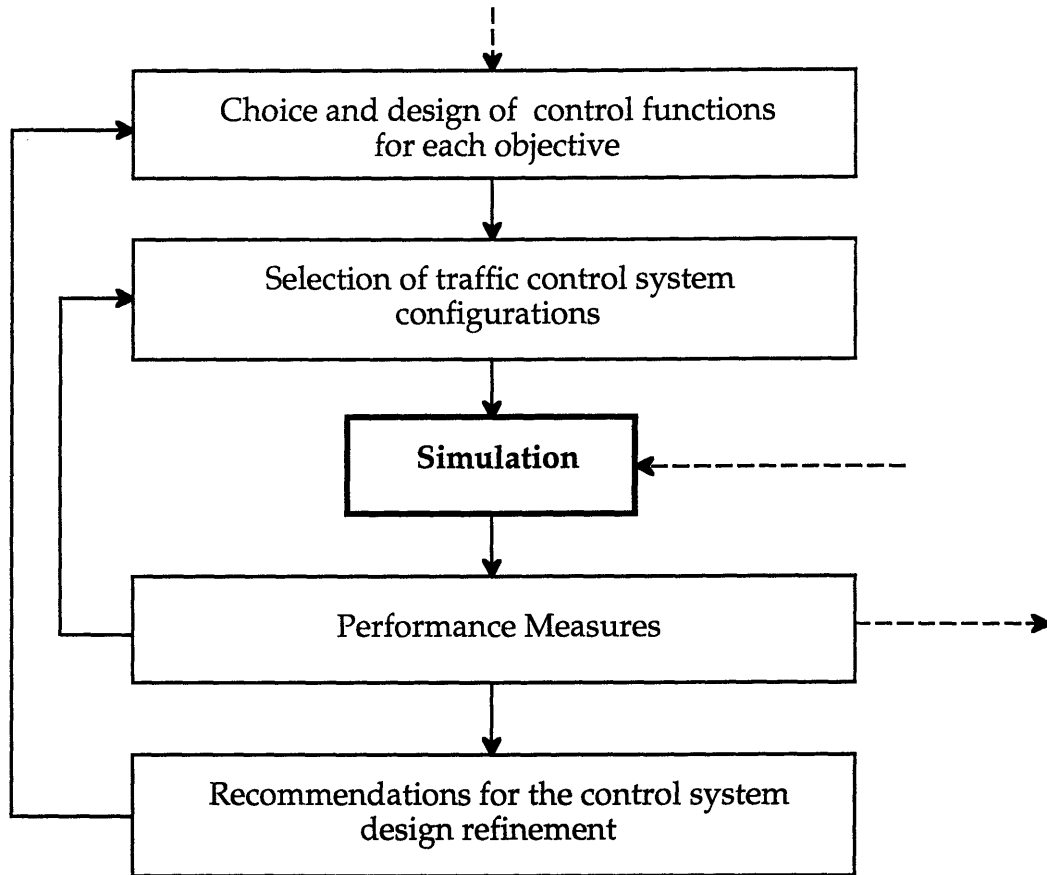
Figure 3-9   The configuration update and design refinement loops

Initially, the different control functions (configuration components) for the objectives are designed. Configurations obtained as combinations of these control function designs define the traffic control strategies for use in the simulation. Each of the configurations is tested against a broad range of scenarios (for more details, refer to section 3.4). The configurations may contain one or more components. Hence, they may or may not cater to all the objectives of the control system. The configurations that cater to all the objectives are referred to as control system designs. Various configurations are tested in the inner loop.

As mentioned in section 3.3, the inner loop is visited a number of times for each iteration of the outer loop. Initial iteration of the inner loop involves simulating configurations with single or two components. This assists in the understanding of the effect of each component on the traffic flow. In addition, the interaction between the

different components can be examined. The results from this iteration are used in choosing the combinations of components to be tested in subsequent iterations. If a combination of components provides higher benefits than either of its individual components, it implies that the components aid each other in alleviating the congestion. On the other hand, if the benefits are lower than either of its individual components, it implies that the components work against each other. These combinations are not further used in the inner loop until the flaws are identified and appropriate refinements are made to the designs in the outer loop. After a sufficient understanding of the interactions is gained, one last iteration of the inner loop is conducted with appropriate configurations that attempt to achieve all the objectives (control system designs). The performance measures from the simulation of these designs are analyzed using multiobjective evaluation and the "best" set of designs (or a "best" design, if only one is chosen) are extracted.

Once all iterations of the inner loop are completed, the results are used to provide recommendations for design refinements. Accordingly, changes are made to the control function designs (through modification of algorithms) including the communication between them. Stress is laid on the utilization of information on the performance of each component and the interaction between components. The next iteration of the outer loop is executed traversing the same steps. Since the control function designs are altered, the interaction between components changes. Hence, the information on the interactions obtained in the previous iterations is invalid. However, the combinations in which none of the components undergo any changes in the designs need not be repeated. This iteration of the outer loop produces a new set of best designs.

Thus, each visit to the outer loop provides a best set of control system designs. At the end of every iteration, all the "best" designs found upto that point are analyzed using multiobjective evaluation to produce a modified set of best designs for the subsequent iterations. This process is continued until an effective, efficient, and robust design is obtained.

93

4 Case Study Design

4.1 Objectives of the Case Study

The evaluation and the design refinement of a control system, as outlined in this thesis, is a complex process. A small scale demonstration of the methodology is planned as a next step. The demonstration is planned on a small network, with some assumptions regarding the control systems and the scenarios. The control systems are generated with a subset of the goals, objectives, and control functions identified in the thesis. Also, the scenarios are prespecified. Hence, the case study designed represents only the first iteration of the iterative processes involving the control design and scenario generation. For this case study, a portion of the Central Artery/Tunnel (CA/T) project of Boston has been identified. Section 4.2 provides a background on the CA/T project. Section 4.3 describes the design of the case study.

4.2 Background on Central Artery/Third Harbor Tunnel (CA/T)

The CA/T project represents the last significant expansion of highway capacity in urban Boston for the foreseeable future. Expanding highway capacity can occur in three ways:

- new, expanded construction of a highway
- utilizing "smart highway technology", and
- improved vehicle capacity (public transportation and HOV lanes)

The CA/T project employs all of these methods [Pardo 93].

4.2.1 Overview of the CA/T Project

The Central Artery/Third Harbor Tunnel project is a 7.5 mile interstate highway project (approximately half of which is tunnel) that provides for the replacement of the elevated Central Artery (I-93) and includes an enlarged, subsurface expressway, and a new tunnel beneath the inner harbor to improve access to Logan Airport. When completed, it will be the largest, most complex underground highway system in the

94

world. It will also be one of the busiest, 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 [Hotz 94, Pardo 93]. The project is designed to double existing traffic capacity.

The highway system will be opened in phases, over a seven year period. The first phase opens the Third Harbor Tunnel segment which is a two-way, four lane, controlled access toll highway, approximately 4 miles in length and connects to the Southeast Expressway (I-93), and South Station on the south side of the harbor, with Logan Airport and Route 1A on the north side of the harbor. The connection of the harbor tunnel to the Massachusetts Turnpike (I-90) will open in a later phase. Later phases also replace the present elevated viaduct comprising I-93 with a new 8-10 lane widened underground freeway. In total, the CA/T network will comprise 107 lane-miles of roadway, 37 lane miles of which will be in covered sections or submerged-tube tunnels. Figure 4-1 shows the Central/Artery project with its underground elements.
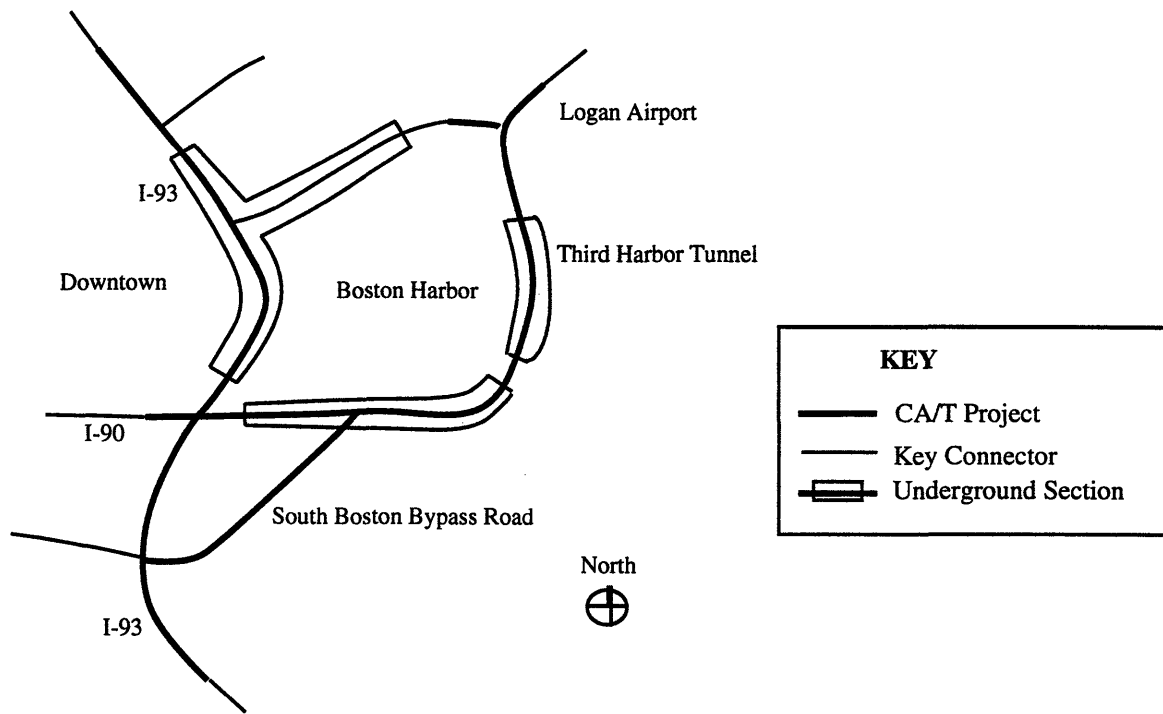

Figure 4-1 The Central Artery/Tunnel project showing underground elements
Source: [Pardo 93]

4.2.2 Advanced Highway Control on the Central Artery/Tunnel

Building a management system for a network as complicated as CA/T poses many challenges. The problem is compounded by the presence of tunnel sections throughout the network. A safe and efficient operation of the highway and tunnels of the Boston Central Artery/Tunnel Project system is achieved using an advanced highway management system. With an integrated control system at its center, the management system incorporates the following significant capabilities:

- *Fully Integrated Operations Control.* A state-of-the-art advanced traffic management system called IPCS (for Integrated Project Control System) which integrates traffic surveillance, incident detection, roadway control, facilities control, and fire and security systems.

- *Life and Safety Monitoring.* The system continually monitors CO (for Carbon Monoxide) levels, fire in the tunnel and supporting buildings, and hydrocarbons in storm water drainage to assure the roadway is safe.

- *Complete Radio Frequency Retransmission Capability.* This capability provides police, fire, and emergency rescue service communication throughout the tunnel, as well as AM/FM radio broadcast.

- *Electronic Toll System.* An advanced electronic toll system for the new harbor tunnel will permit cars to drive through and pay their tolls without slowing down.

- *Overheight Vehicle Control.* The system checks all vehicle heights at entry points. An overheight vehicle attempting to enter the system is given warnings and signals are used to control its access. In the worst case, physical barriers are used to prevent its entry.

- *Route Guidance.* The system allows strategic diversion of traffic to prevent a minor breakdown or fender-bender from becoming a major headache. This tactic is used for events related to special conditions also (e.g., Thanksgiving airport rush, a Red Sox game).

4.3 Case Study Design

This subsection presents the details of the case study design. The discussion is divided into two parts, which together describe the inputs to the simulation. The first part describes the different control configurations that will be tested, and the second part, the scenarios that define the traffic conditions to test the control configurations. The

configurations enumerated represent the configurations tested only in the first iteration of the inner and outer loops (Figures 3-1 and 3-2). Also, the scenarios specified cover only the first iteration of the scenario generation loop (Figure 3-6). Note that the scenario generation method presented in subsection 3.4.2 involves a single event pattern for the first set of scenarios. Since in this case, no iterations will be conducted, more than one event pattern is specified. Performance measures relevant to the goals of the designed control system are selected from those enumerated in subsection 3.5.1 and Appendix A. The MIcroscopic Traffic SIMulator (MITSIM) developed at MIT is used as the simulation tool. For a detailed description of the model, refer to [Yang 93].

4.3.1 Control Configurations

For demonstration of the methodology, two of the goals identified in Chapter 3 are considered.
- Improve level of service
- Reduce CO emissions and exposure

A total of three objectives from among those identified for these goals are considered. The three objectives are:
- Reduce travel time (combines travel time by link and OD pair, for purposes of simplification).
- Reduce time spent in queue.
- Reduce link density.

As a reminder, the control functions that could be used independently or in combinations to achieve each of these objectives are shown in Table 4-1.

Table 4-1 Control functions for objectives of case study

Objective	Route Diversion	Access Control	Mainline Control	Arterial Control
Reduce travel time	√	√	√	√
Reduce time spent in queue	√	√	√	√
Reduce link density	√	√	√	√

In this case study design, 'Route Diversion' and 'Arterial Control' are not implemented. Therefore, the two control functions 'Mainline Control' and 'Access Control' are used in designing the strategies for achieving the objectives. The various

97

control devices used for each of these two control functions for the case study are as follows:

- Lane Use Signs, Variable Speed Limit Signs and Mainline Meters for Mainline Control
- Ramp Meters and Portal Signals for Access Control.

Although both the control functions chosen for implementation could be used for each of the chosen objectives, only one control function for each objective is chosen for the case study. This may not be the best means of achieving the objectives, however, they will suffice for demonstration purposes. The control functions chosen for the objectives are:

- Reduce travel time: Access Control
- Reduce time spent in queue: Mainline Control
- Reduce link density: Mainline Control

Therefore, the control components for the case study are:

- Access Control for reducing travel time (A)
- Mainline Control for reducing time spent in queue (B)
- Mainline Control for reducing link density (C)

Given these, a total of seven control configurations are possible:

- A
- B
- C
- A + B
- A + C
- B + C
- A + B + C

The analysis of the results from the usage of single component configurations will give the individual strength of the configuration in achieving the objective. The double and triple component configurations help in explaining the interaction effects among the different components (how the results are affected when a component is used in conjunction with one or more other components). All the seven control configurations are used in the case study. Simulations will be executed using each of these configurations and the recommendations to the design (not presented in this thesis) made based on the resulting performance measures.

Appendix C describes the algorithms designed for each of the control functions. The algorithms described represent only a first cut and need to be refined and debugged before implementation in this case study.

4.3.2 Scenarios

The network used for the simulation is a part of the overall Central Artery/Tunnel network comprising of a portion of Interstate 90, the South Boston Bypass Road, the South Boston Interchange, Third Harbor Tunnel, and Logan Airport Interchange (see Figure 4-1). In the case study, all the links (except one), are assumed to be tunnel sections with zero grade. Also, the network is simplified as follows:

- Reducing the South Boston Interchange to a single internal node[1] and making the South Boston Bypass Road an on-ramp to the tunnel

- Reducing the Logan Airport Interchange to a single internal node and making the exit to Logan airport from Interstate 90 East bound an off-ramp to the tunnel

- Considering only unidirectional flow for simulation: East bound flow is considered

The simplified network is shown in Figure 4-2.
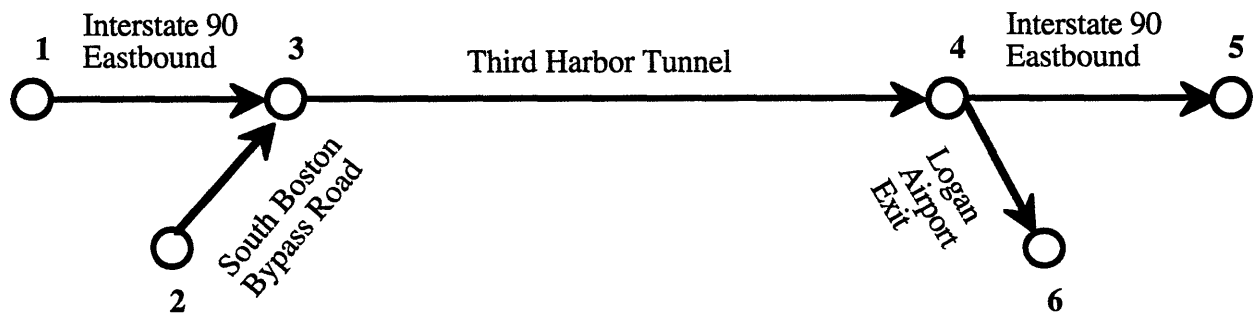

Figure 4-2 Simplified Experimental Network

---

[1] The origins and destinations of vehicles are considered to be external nodes and the intermediate nodes that lie on the vehicle's route are referred to as internal nodes.

The network consists of the following
- 6 nodes (1, 2, 3, 4, 5, 6): 1,2,5,6 are external nodes and 3,4 are internal nodes
- 5 links

      (1,3): Link 0

      (2,3): Link 1

      (3,4): Link 2

      (4,5): Link 3 and

      (4,6): Link 4

The links (1,3), (3,4) and (4,5) form the mainline and are covered sections. The link (2,3) is an on-ramp and is an open section. The link (4,6) is an off-ramp and is a covered section.

Links are divided into segments based on the differences in the geometric characteristics of different sections of the link. A link with the same geometric characteristics may be divided into segments if its length exceeds a certain limit.

Three types of vehicles are considered.
- Single Occupancy Vehicles (New Cars + Old Cars)
- High Occupancy Vehicles (Buses)
- Trucks (Single-unit Trucks and Trailer Trucks)

Section 3.4 gave a description of the three factors - traffic demand, events, driver behavior - in terms of which scenarios are captured. Two demand patterns capturing normal and congested conditions relative to the peak traffic conditions are specified. Two event patterns, each with two events are considered (again, two event patterns instead of one are specified because no iterations will be conducted in this case study). Similarly, two sets of behavioral parameters representing the nominal and aggressive behavior of drivers are specified. Thus, a total of 8 scenarios are possible as combinations of the three factors. A detailed description of the demand patterns, event patterns and behavioral parameter sets that are used in generating the 8 scenarios is given in Appendix D.

5 Conclusion and Future Research

5.1 Conclusion

Technological advances in electronic control and communication have facilitated the design of dynamic traffic control systems. While a number of such control systems are already in operation all around the world and are undergoing an operational evaluation process, more systems are being designed. Computer models that realistically simulate the various components of the traffic control system are gaining wide applicability to the testing of candidate control strategies. The objective of this research work is to utilize the concepts of microsimulation models to evaluate traffic control systems before operational tests are conducted. This allows for testing many designs and choosing a best one. Evaluation is an integral part of the design process due to the complexity of the relationship between the control systems and the traffic flows. Hence, a methodology for the design and evaluation of real-time traffic control systems that integrate the various aspects of ATMS and ATIS concepts of IVHS is developed.

The approach taken utilizes the concept of gradually improving the design through iterative improvements. The design task is broken down into smaller tasks. Initially, the various goals and objectives of the integrated traffic control system are identified and various control functions (or strategies) are designed to achieve each of the objectives. The advantages of each of these control function designs and the type of interaction with each other are understood through the analysis of performance measures obtained through simulations under a wide spectrum of scenarios. This knowledge is utilized in integrating the individual control function designs to obtain a comprehensive design that addresses all objectives.

The methodology described in this thesis can be used to design and evaluate any traffic management system using simulation models.

5.2 Future Research

An important part of the development of a methodology is its demonstration. A case study design was presented in Chapter 4 for a small scale demonstration of the methodology. In this section, the various aspects of the full scale implementation and demonstration of the methodology are presented. The issues fall under five major categories:

- Traffic network
- Surveillance system
- Control system
- Scenarios
- Performance Measures

The specifics of each are described below.

With regards to the *traffic network*, the various components of the traffic network need to be simulated along with the vehicle movements. Similarly, the various components of the *surveillance systems* need to be realistically simulated. With regards to the *control system* design and evaluation, the following needs to be done.

- Identification and development of designs for each of the control functions used for the objectives.
- Implementation of these designs as responses in the traffic simulation.
- Confirmations of the a priori expectations on the type of interaction between the objectives.
- The objectives CO(VFS), SV(TTE), SV(PV), and CS(SS) along with their interactions with other objectives should be properly understood through simulation to aid in the design process.
- Once the exact interaction effects are recognized, the weak interactions should also be specified in terms of positive or negative effects.
- In the interaction table presented in this thesis, the cells which have the same entries do not necessarily signify the same amount of interaction. The interaction between two objectives needs to be quantified, instead of stating it qualitatively, as is done in the thesis. This will further assist in the design process. This can be done using a rating scheme. For example, assuming a rating scheme of 1 to 5,

    5- is the strongest conflict between two objectives

    1- is the weakest conflict between two objectives

0 implies absence of interaction between two objectives

1+ is the weakest positive interaction between two objectives

5+ is the strongest positive interaction between two objectives.

With respect to the *scenarios*, the thesis only specified the various inputs to generate scenarios. A lot of information is necessary before scenarios could be formulated.

- The various traffic demand patterns need to be obtained and organized as origin-destination patterns for input to the simulator.

- Data corresponding to the type of incidents, frequency of each type, their severity in terms of durations (means and standard deviations), their probability of occurrence should be obtained to be able to represent incidents realistically.

- Sensitivity analyses are performed to determine exactly the effect of the behavioral parameters on the performance measures. This assists in the choice of the behavioral parameters sets for use in the simulation.

A post processor is needed for the simulator that presents the various *performance measures* identified in this thesis from the crude outputs of the simulation.

The methodology presented in this thesis is still skeletal in nature. Further development and refinement is necessary. The results of implementing the case study design presented in Chapter 4 should assist this methodology development and refinement effort.

Appendix A Specification of Performance Measures

In this appendix, the specific measures that are derived from each class of measures are described. Mathematical expressions are presented that help in deriving these measures from the crude output of the simulation. In each class description, the structure followed is that of initially defining the elemental observations, introducing the notation and finally specifying the measure. An elemental observation is the basic measure by which the measures are computed.

A.1 Carbon Monoxide Emissions and Exposure

CO Emissions

The *elemental observations* are the CO emitted by each vehicle over its journey.

Let C_n^e = the CO emitted by vehicle n.

         N = the total number of vehicles.

The following measures are obtained.

• Total CO emissions across all the vehicles in the network, given by:

$$C^e = \sum_{n=1}^{N} C_n^e \qquad\qquad (A.1)$$

CO Exposure

The CO level measurements are computed for each vehicle every minute. The *elemental observations* are the cumulative CO exposures of vehicles obtained as the sum of the CO exposures in each of the time intervals.

Let N = the total number of vehicles.

M_n = the total number of CO exposure observations of vehicle n.

c_{mn} = the mth CO level observation encountered by vehicle n.

C_n = the cumulative CO exposure of the vehicle n over its entire journey.

The elemental cumulative CO exposure of a vehicle n is given by:

$$C_n = \sum_{m=1}^{M_n} c_{mn} \qquad \text{(A.2)}$$

The measures computed based on these elemental observations are:

- Maximum cumulative exposure, C^{max}. This measure gives an indication of the worst effect. The maximum exposure is given by:

$$C^{max} = \underset{n}{\text{Max}} \ (C_n) \qquad \text{(A.3)}$$

- Mean cumulative exposure computed over all the vehicles, $\overline{C}$. The mean exposure gives a general idea of the effect of a control strategy and a scenario on CO levels. The mean is given by:

$$\overline{C} = \frac{1}{N} \sum_{n=1}^{N} C_n \qquad \text{(A.4)}$$

- Standard deviation of cumulative exposure computed over all the vehicles, s_c. The standard deviation provides a range of exposure within which most of the vehicles lie. The standard deviation is given by:

$$s_c = \sqrt{\frac{\sum_{n=1}^{N} (C_n - \overline{C})^2}{N - 1}} \qquad \text{(A.5)}$$

- Median of cumulative exposure over all the vehicles, C^{median}, given by:

$$C^{median} = \underset{n}{\text{Median}} \ (C_n) \qquad \text{(A.6)}$$

Severity of CO exposure

While the above measures capture exposure, the following measures capture the severity of the exposure. They involve the elemental observations whose values exceed the safety limit.

- Percentage of vehicles exposed above the safety limit, P_c.

Let $C^* =$ safety limit

$$I_n = \begin{cases} 1 & \text{if } C_n > C^* \\ 0 & \text{otherwise} \end{cases} \tag{A.7}$$

The percentage is then given by:

$$P_c = \frac{100}{N} \sum_{n=1}^{N} I_n \tag{A.8}$$

- The mean exposure for vehicles that exceed the limit, $\overline{C}_{C_n > C^*}$, is given by:

$$\overline{C}_{C_n > C^*} = \frac{\sum_{n=1}^{N} C_n I_n}{\sum_{n=1}^{N} I_n} \tag{A.9}$$

- The mean excess exposure relative to safety limit, $P_{\overline{C}_{C_n > C^*}}$. This measure indicates the amount by which the conditional mean exceeds the safety limit and is given by:

$$P_{\overline{C}_{C_n > C^*}} = 100 \text{ x } \frac{\overline{C}_{C_n > C^*} - C^*}{C^*} \tag{A.10}$$

- The maximum excess exposure relative to safety limit, $P_{C^{max}}$. his measure yields the distance between the maximum exposure over all vehicles and the safety limit and is given by:

$$P_{C^{max}} = 100 \text{ x } \frac{C^{max} - C^*}{C^*} \tag{A.11}$$

• Sensitivity of percentage above the limit with respect to the limit. It is of interest to examine how a change in the safety limit will alter the percentage of vehicles exceeding the specified exposure limits. The measure is the change in the percentage of vehicles above the safety limit for a unit change in the safety limit. This represents the slope of the cumulative distribution function (cdf) curve at the safety limit. To show how the cdf can be used, two extreme situations that are possible are illustrated in Figures A-1a and A-1b.

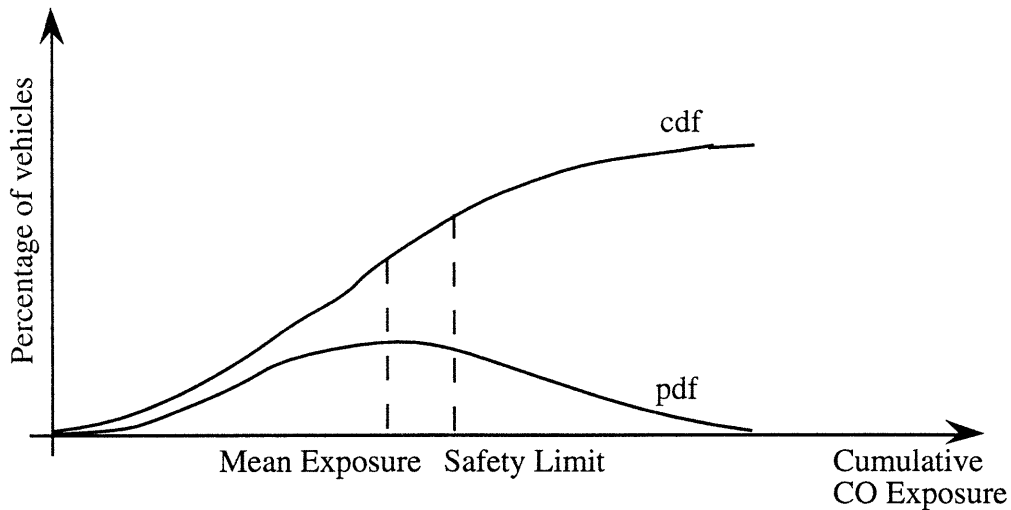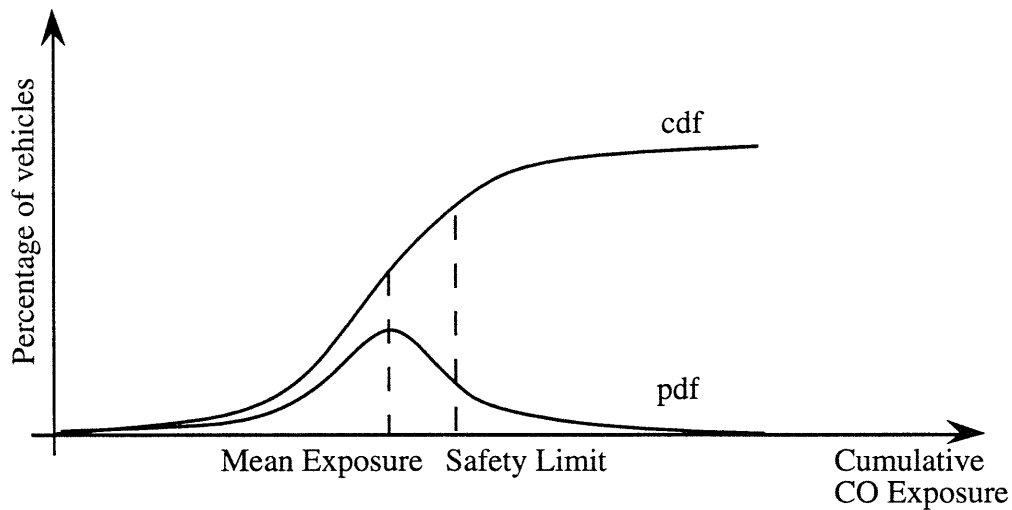**Figure A-1a** Flat distribution of cumulative CO exposure over vehicles

Figure A-1b Sharp distribution of cumulative CO exposure over vehicles

In the first case, different vehicles are subjected to a wide range of CO exposures. Hence, for a small change in the safety limit, the percentage of vehicles above the limit will not change substantially. However, in the latter case, we have a narrow pdf. In this

107

case, a small change in the safety limit will result in a significant change in the percentage of vehicles above the limit.

Let P_1 be the percentage of vehicles above the safety limit C_1^*. Let this change to P_2 when the safety limit changes to C_2^*. Then, the change in the percentage of vehicles above the safety limit for a unit change in the safety limit is given by:

$$\Delta P_{C^{max}} = \frac{P_2 - P_1}{C_2^* - C_1^*}$$

(A-12)

where C_2^* is chosen appropriately.

A.2 Travel Time

Link Travel Time

The *elemental observations* are the ratio of the average travel times obtained by aggregating all vehicles entering the link during 5 minute intervals to the free-flow travel time.

Let m = the entering time interval ($m = 1$ to M).

P = the total number of links ($p = 1$ to P).

N_{pm} = the total number of vehicles of entering link p during time interval m.

t_{npm} = the time for vehicle n traverse link p entering during time interval m.

f_p = the free-flow travel time of link p.

The average travel time for all vehicles that enter link p during time interval m is given by:

$$\bar{t}_{pm} = \frac{\sum_{m=1}^{N_{pm}} t_{npm}}{N_{pm}}$$

(A.13)

The elemental observation — the ratio of the average travel time for all vehicles that enter link p during time interval m to the free-flow travel time on the link — is given by:

$$r_{pm} = \frac{\bar{t}_{pm}}{f_p}$$

The following are the criteria for selecting the links:

- The link p^{max} associated with the maximum ratio over all time intervals and over all links is identified. The maximum is given by:

$$r_{p^{max}}^{max} = \underset{p,m}{\text{Max}} \, (r_{pm}) \qquad (A.14)$$

The following measures are computed for the selected links.

• Mean ratio over all time intervals, given by:

$$\bar{r}_{p^{max}} = \frac{\sum_{m=1}^{M} r_{p^{max}m}}{M} \qquad (A.15)$$

• Standard deviation of the ratio over all the time intervals.

$$s_{p^{max}} = \sqrt{\frac{\sum_{m=1}^{M} (r_{p^{max}m} - \bar{r}_{p^{max}})}{M-1}} \qquad (A.16)$$

Origin-Destination Travel Time

The *elemental observations* for the calculation of the measures are the average travel times obtained by aggregating all vehicles of a particular category departing from an origin to a particular destination during 5 minute intervals.

Let   O and D be the origin and destination of the vehicle under consideration.

v = vehicle category (v = 1 to V).

m = the departing time interval (m = 1 to M).

p = the O-D pair (p = 1 to P).

N_{vpm} = the total number of vehicles of category v traveling on O-D pair p during time interval m.

t_{nvpm} = the time for vehicle n of category v to travel on O-D p departing during time interval m.

The elemental observation — the average travel time for all vehicles of category v that departed origin O during time interval m and traveling towards destination D — is given by:

$$\bar{t}_{vpm} = \frac{1}{N_{vpm}} \sum_{n=1}^{N_{vpm}} t_{nvpm} \qquad \text{(A.17)}$$

With respect to the aggregation given in Table 3-5, M can be the number of time intervals in an extended period or a peak period and P can be the number of O-D pairs in the whole network or the O-D pairs of special interest.

The following measures are obtained:

• Mean travel time, $\bar{t}$. The average travel time for vehicles of category v belonging to O-D pair p across the time intervals of interest is given by:

$$\bar{t}_{vp} = \frac{1}{M} \sum_{m=1}^{M} \bar{t}_{vpm} \qquad \text{(A.18)}$$

The average travel time for vehicles of category v across a set of O-D pairs and across the time intervals of interest is given by:

$$\bar{t}_{v} = \frac{1}{PM} \sum_{p=1}^{P} \sum_{m=1}^{M} \bar{t}_{vpm} \qquad \text{(A.19)}$$

The average travel time for a particular O-D pair p across all vehicle categories and across the time intervals of interest is given by:

$$\bar{t}_{p} = \frac{1}{VM} \sum_{v=1}^{V} \sum_{m=1}^{M} \bar{t}_{vpm} \qquad \text{(A.20)}$$

The average travel time across all the vehicle categories, a set of O-D pairs, and the time intervals of interest is given by:

$$\bar{t} = \frac{1}{VPM} \sum_{v=1}^{V} \sum_{p=1}^{P} \sum_{m=1}^{M} \bar{t}_{vpm} \qquad \text{(A.21)}$$

110

- Mean 'travel time to free-flow travel time ratio'.

Let f_p represent the free-flow travel time for the O-D pair p. The free-flow travel time for an O-D pair is the average of the free-flow travel times on the different paths (if there exists more than one) of that O-D pair. The free-flow travel time is defined as the travel time when vehicles are traveling at the design speed limits. The ratio of the actual to free-flow travel time for vehicle n of category v belonging to O-D pair p during time interval m is given by:

$$r_{vpm} = \frac{\bar{t}_{vpm}}{f_p} \qquad (A.22)$$

The measures representing the various aggregation schemes are computed in a similar fashion as the mean travel time measures presented above. The ratio r_{vpm} is used instead of $\bar{t}_{vpm}$.

- Maximum travel time is given by:

$$\bar{t}^{max} = \text{Max} \ (\bar{t}_{vpm}) \qquad (A.23)$$

This is computed across the appropriate variables v, p, m for the various aggregation schemes.

- Minimum travel time is given by:

$$\bar{t}^{min} = \text{Min} \ (\bar{t}_{vpm}) \qquad (A.24)$$

This is computed across the appropriate variables v, p, m for the various aggregation schemes.

- Standard Deviation, s is computed across the appropriate variables v, p, m for the various aggregation schemes.

- Median travel time, $\bar{t}^{median}$, is given by
$$\bar{t}^{median} = \text{Median} \ (\bar{t}_{vpm}) \qquad (A.25)$$

This is computed across the appropriate variables v, p, m for the various aggregation schemes.

Travel Time Imbalances Across Paths

The *elemental observation*, r_{vpm}, is the ratio of the imbalance to the mean O-D travel time (for an O-D pair, a vehicle category, and a departure time interval).

Let
j_p = the jth path of O-D pair p.

J_p = the total number of alternate paths of O-D pair p (j_p = 1 to J_p).

t_{nvpmj_p} = the travel time of the nth vehicle of category v, associated with O-D pair

p, departing during time interval m, and traveling on path j.

N_{vpmj_p} = the total number of vehicles of category v, associated with O-D pair

p, departing during time interval m, and traveling on path j.

The average travel time of all vehicles of category v that have departed during the time interval m from origin O towards destination D (path j of O-D pair p) is given by:

$$\bar{t}_{vpmj_p} = \frac{1}{N_{vpmj_p}} \sum_{n=1}^{N_{vpmj_p}} t_{nvpmj_p} \tag{A.26}$$

And, consequently, the elemental observation is given by:

$$r_{vpm} = \frac{\underset{j_p}{\text{Max}} (\bar{t}_{vpmj_p}) - \underset{j_p}{\text{Min}} (\bar{t}_{vpmj_p})}{\bar{t}_{vpm}} \tag{A.27}$$

The following measures are obtained.

• Maximum 'imbalance to mean O-D travel time ratio', given by:

$$r^{max} = \text{Max} (r_{vpm}) \tag{A.27}$$

This is computed across the appropriate variables v, p, and m for the various aggregation schemes.

A.3 Speed

Speeds at Point Locations

The *elemental observations* are the time-mean-speeds at specific locations for short intervals of time (in the order of half a minute).

Let $p = 1, \ldots, P$, be the locations chosen to be monitored.

      $m = 1, \ldots, M$, be the time intervals.

      N_{mp} = the total number of vehicles crossing point p during time interval m.

      u_{npm} = the speed of vehicle n crossing point p during time interval m.

      $\bar{u}^t_{pm}$ = the time-mean-speed at point p for time interval m.

The time-mean-speed at point p for time interval m is given by:

$$\bar{u}^t_{pm} = \frac{1}{N_{mp}} \sum_{n=1}^{N_{mp}} u_{npm} \qquad (A.28)$$

The following measures are computed for the selected locations.

- Mean of time-mean-speed over time for the entire simulation period, given by:

$$\bar{\bar{u}}^t_p = \frac{1}{M} \sum_{m=1}^{M} \bar{u}^t_{pm} \qquad (A.29)$$

- Standard deviation of time-mean-speed over time for the entire simulation period, given by:

$$s^t_p = \sqrt{\frac{\sum_{m=1}^{M} (\bar{u}^t_{pm} - \bar{\bar{u}}^t_p)^2}{M-1}} \qquad (A.30)$$

- First order autocorrelation coefficient of time-mean-speed over time, ρ^t_p. This measure captures the nature of the dependence of the current speed on the speed from the previous time period. A high positive correlation indicates a smooth variation of speeds over consequent time intervals and hence a stable control system. A high negative correlation

indicates an abrupt variation of speeds over consequent time intervals which may be indicative of an unstable control system. The first order autocorrelation coefficient is given by:

$$\rho_p^t = \frac{\displaystyle\sum_{m=2}^{M}(\overline{u}_{pm}^t - \overline{u}_0)(\overline{u}_{p(m-1)}^t - \overline{u}_{-1})}{\sqrt{\displaystyle\sum_{m=2}^{M}(\overline{u}_{pm}^t - \overline{u}_0)^2 \sum_{m=2}^{M}(\overline{u}_{p(m-1)}^t - \overline{u}_{-1})^2}} \qquad (A.31)$$

where

$$\overline{u}_0 = \frac{\displaystyle\sum_{m=2}^{M}\overline{u}_{pm}^t}{M-1} \quad \text{and} \quad \overline{u}_{-1} = \frac{\displaystyle\sum_{m=1}^{M-1}\overline{u}_{pm}^t}{M-1} \qquad (A.32)$$

Traffic Speed - Speed Limit Differentials

Let M = the total number of time points.

Z = the total number of sections.

$\overline{u}_{zm}^s$ = the space-mean-speed for section z at time point m.

u_{zm}^* = the posted speed limit for section z at time point m.

δ_{zm} = the traffic speed-speed limit differential for section z at time point m.

t_{nzm} = the time vehicle n spends traversing section z in association with time point m.

y_z = the length of the section z.

N_{zm} = the total number of vehicles on section z at time point m.

The average speed of vehicle n in traversing section z in association with time point m, $\overline{u}_{nzm}$ is given by:

$$\overline{u}_{nzm} = \frac{y_z}{t_{nzm}} \qquad (A.33)$$

The space-mean-speed for time point m at section z is given by:

$$\overline{u}_{zm}^s = \sum_{n=1}^{N_{zm}} \overline{u}_{nzm} \frac{t_{nzm}}{\displaystyle\sum_{n=1}^{N_{zm}} t_{nzm}} = \frac{N_{zm} y_z}{\displaystyle\sum_{n=1}^{N_{zm}} t_{nzm}} \qquad (A.34)$$

114

And, consequently, the elemental observation — traffic speed - speed limit differential for section z at time point m — is given by:

$$\delta_{zm} = \bar{u}^s_{zm} - u^*_{zm} \qquad (A.35)$$

The measures calculated for each section as follows:

- Maximum differential over time for section z, given by:

$$\delta^{max}_z = \underset{m}{\text{Max}}\,(\delta_{zm}) \qquad (A.36)$$

- Mean differential over time for section z, given by:

$$\bar{\delta}_z = \frac{1}{M}\sum_{m=1}^{M}\delta_{zm} \qquad (A.37)$$

- Standard deviation of differential over time for section z, given by:

$$s_\delta = \sqrt{\frac{\sum_{m=1}^{M}(\delta_{zm} - \bar{\delta}_z)^2}{M}} \qquad (A.38)$$

Speed During Journey

The measures capture the average, standard deviation, maximum, minimum and the temporal first-order autocorrelation for specific vehicles that exhibit the maximum mean and the maximum standard deviation of speed and for a "typical" vehicle. i.e., one that exhibits the median of mean speed. In addition, they capture the average and standard deviation of speed over all vehicles.

The *elemental observations* associated with the measures are the speeds of each vehicle measured at equally spaced time points (e.g. every 1 minute) during the vehicle's entire journey. Once a vehicle has completed its journey, its mean speed and standard deviation over the entire journey are calculated.

Let N = the total number of vehicles.

M = the total number of time points.

u_{mn} = the speed of vehicle n for observation m.

Then, the mean speed of vehicle n over its entire journey is given by:

$$\bar{u}_n = \frac{1}{M} \sum_{m=1}^{M} u_{mn}$$ (A.39)

The standard deviation of a vehicle n over its entire journey is given by:

$$s_{u_n} = \sqrt{\frac{\sum_{m=1}^{M} (u_{mn} - \bar{u}_n)^2}{M-1}}$$ (A.40)

The maximum of the mean over all vehicles is given by:

$$\bar{u}_{n_{\bar{u}}^{max}}^{max} = \underset{n}{\text{Max}}\,(\bar{u}_n)$$ (A.41)

where $n_{\bar{u}}^{max}$ identifies the vehicle with the maximum mean speed.

Similarly, the median of the mean over all vehicles is given by:

$$\bar{u}_{n_{\bar{u}}^{median}}^{median} = \underset{n}{\text{Median}}\,(\bar{u}_n)$$ (A.42)

where $n_{\bar{u}}^{median}$ identifies the vehicle with the median of mean speed.

The maximum of the standard deviation is given by:

$$s_{n_s^{max}}^{max} = \underset{n}{\text{Max}}\,(s_{u_n})$$ (A.43)

where n_s^{max} identifies the vehicle with the maximum standard deviation.

Then, the measures are computed as follows:

• Mean speed for vehicle, $n_{\bar{u}}^{max}$, as defined in A.39.

• Standard deviation of speed over time for vehicle, $n_{\bar{u}}^{max}$, as defined in A.40.

• Maximum speed of vehicle $n_{\bar{u}}^{max}$ given by:

$$u_{n_{\bar{u}}^{max}}^{max} = \underset{m}{\text{Max}} \left(u_{mn_{\bar{u}}^{max}} \right) \tag{A.44}$$

- Minimum speed of the vehicle $n_{\bar{u}}^{max}$ given by:

$$u_{n_{\bar{u}}^{max}}^{min} = \underset{m}{\text{Min}} \left(u_{mn_{\bar{u}}^{max}} \right) \tag{A.45}$$

- Temporal first order autocorrelation of speed for vehicle $n_{\bar{u}}^{max}$ given by:

$$\rho_{n_{\bar{u}}^{max}}^{u} = \frac{\displaystyle\sum_{m=2}^{M} (u_{mn_{\bar{u}}^{max}} - \bar{u}_0)(u_{(m-1)n_{\bar{u}}^{max}} - \bar{u}_{-1})}{\sqrt{\displaystyle\sum_{m=2}^{M} (u_{mn_{\bar{u}}^{max}} - \bar{u}_0)^2 \sum_{m=2}^{M} (u_{(m-1)n_{\bar{u}}^{max}} - \bar{u}_{-1})^2}} \tag{A.46}$$

where

$$\bar{u}_0 = \frac{\displaystyle\sum_{m=2}^{M} u_{mn_{\bar{u}}^{max}}}{M-1} \quad \text{and} \quad \bar{u}_{-1} = \frac{\displaystyle\sum_{m=1}^{M-1} u_{mn_{\bar{u}}^{max}}}{M-1} \tag{A.47}$$

Similarly, all the above measures are computed for vehicles $n_{\bar{u}}^{median}$ and n_{S}^{max}.

- Mean of mean speeds over all vehicles given by:

$$\bar{u} = \frac{1}{N} \sum_{n=1}^{N} \bar{u}_n \tag{A.48}$$

- Standard deviation of mean speeds for all vehicles given by:

$$s_{\bar{u}} = \sqrt{\frac{\displaystyle\sum_{n=1}^{N} (\bar{u}_n - \bar{u})^2}{N-1}} \tag{A.49}$$

A.4 Lane Changing

Lane Changing during journey

The performance measures capture the average and standard deviation of the number of lane changes over all vehicles and for the specific vehicles that exhibits the maximum mean number of lane changes and median of mean number of lane changes (i.e., the typical vehicle).

The *elemental observations* associated with the measures are the number of lane changes per minute for each vehicle computed over 5 minute intervals.

Let N = the total number of vehicles.

M_n = the total number of intervals associated with vehicle n.

l_{mn} = the number of lane changes per minute of vehicle n for time interval m.

The mean number of lane changes per minute for vehicle n over its journey is given by:

$$\bar{l}_n = \frac{1}{M_n} \sum_{m=1}^{M_n} l_{mn} \qquad (A.50)$$

The maximum of the means is given by:

$$\bar{l}_{n_i^{max}}^{max} = \text{Max}_n (\bar{l}_n) \qquad (A.51)$$

where n_i^{max} identifies the vehicle with the maximum mean number of lane changes.

Similarly, the median of the means is given by:

$$\bar{l}_{n_i^{median}}^{median} = \text{Median}_n (\bar{l}_n) \qquad (A.52)$$

where n_i^{median} identifies the vehicle with the median of mean number of lane changes.

Then, the measures are computed as follows:

• Mean number of lane changes per minute for vehicle n_i^{max} as defined by A.50.

- Standard deviation of number of lane changes per minute over journey time for vehicle n_i^{max} given by:

$$s_{n_i^{max}} = \sqrt{\frac{\sum_{m=1}^{M_{n_i^{max}}}(l_{mn_i^{max}} - \bar{l}_{n_i^{max}})^2}{M_{n_i^{max}} - 1}} \qquad (A.53)$$

- Mean number of lane changes per minute for vehicle n_i^{median} as defined by A.50.

- Standard deviation of number of lane changes per minute over journey time for vehicle n^{median} given by:

$$s_{n_i^{median}} = \sqrt{\frac{\sum_{m=1}^{M_{n_i^{median}}}(l_{mn_i^{median}} - \bar{l}_{n_i^{median}})^2}{M_{n_i^{median}} - 1}} \qquad (A.54)$$

- Mean of mean number of lane changes over all vehicles given by:

$$\bar{l} = \frac{1}{N}\sum_{n=1}^{N}\bar{l}_n \qquad (A.55)$$

- Standard deviation of mean number of lane changes for all vehicles given by:

$$s_{\bar{l}} = \sqrt{\frac{\sum_{n=1}^{N}(\bar{l}_n - \bar{l})^2}{N-1}} \qquad (A.56)$$

These measures are computed independently for each category of vehicle and separately for both mandatory and discretionary lane changes.

Lane Changing at Specific Locations

The *elemental observations* are the number of lane changes per lane-mile per minute on the roadway sections obtained for 5 minute intervals.

Let z = the section at which the lane changes are measured (z = 1 to Z).

        m = the time interval during which the lane changes are counted (m = 1 to M).

        l_{mz} = the number of lane changes per lane-mile per minute for time interval m at section z.

The measures are computed as follows for the selected sections:

- Maximum number of lane changes per lane-mile per minute across the 5 minute intervals for a section z given by:

$$l_z^{max} = \underset{m}{\text{Max}}\ (l_{mz})$$
(A.57)

- Mean number of lane changes per lane-mile per minute across the 5 minute intervals given by:

$$\bar{l}_z = \frac{1}{M}\sum_{m=1}^{M} l_{mz}$$
(A.58)

- Standard deviation of number of lane changes per lane-mile per minute across the 5 minute intervals given by:

$$s_{l_z} = \sqrt{\frac{\sum_{m=1}^{M}(l_{mz} - \bar{l}_z)^2}{M-1}}$$
(A.59)

A.5 Density

The *elemental observations* are the number of vehicles per lane-mile for each link over entire paths computed every 5 minutes.

Let P = the total number of paths (p = 1 to P).

        L_p = the number of links on path p (l_p = 1 to L_p).

        $d_{l_p m}$ = the density of link l of path p for a time point m.

The following are the criteria for selecting the paths at the times of interest:

- The path p^{max} associated with the maximum link density over all time points and all paths is identified. The maximum is given by:

$$d_{p^{max}}^{max} = \underset{p,m,l}{\text{Max}} \, (d_{l_p m})$$

(A.60)

- The path $p^{max-mean}$ associated with the maximum mean density along entire path is identified.

Let z_{l_p} = the length of link l of path p.

Z_p = the total length of path p.

The mean density for path p at time point m is given by:

$$d_{pm} = \sum_{l_p=1}^{L_p} \frac{z_{l_p}}{Z_p} d_{l_p m}$$

(A.61)

The maximum mean is given by:

$$d_{p^{max-mean}}^{max} = \underset{p,m}{\text{Max}} \, (d_{pm})$$

(A.62)

- The path p^{median} associated with the median of mean densities along entire path is identified. The median of mean is given by:

$$d_{p^{median}}^{median} = \underset{p,m}{\text{Median}} \, (d_{pm})$$

(A.63)

- The path p^{max-sd}, associated with the maximum of standard deviations of density along entire path is identified. The standard deviation of density along entire path for interval m and path p is given by:

$$s_{pm} = \sqrt{\sum_{l_p=1}^{L_p} \frac{z_{l_p}}{Z_p} (d_{l_p m} - d_{pm})^2}$$

(A.64)

The maximum is given by:

$$s_{p^{max-sd}}^{max} = \underset{p,m}{\text{Max}} \, (s_{pm})$$

(A.65)

121

The following measures are computed for each of the selected paths at specific time points. Denote these selected paths and time points by h and k respectively.

- Mean density along entire path which is given by d_{hk} as defined by A.61.

- Standard deviation of density along entire path is given by s_{hk} as defined by A.64.

A.6 Queuing

Lengths and durations of queues

The *elemental observations* are the graphs that describe queue length as a function of time. The queue length is recorded from the time it is formed until it completely dissipates. The queue lengths at every observation point in time comprise the points of the graph. The maximum queue length and the duration for each queue are correspondingly the maximum values of the "queue curve" along each axis of the graph.

Let Q = the total number of queues of a particular type (see Table 3-6) that have
formed during the entire simulation period.
L_q = the maximum queue length of queue q measured by number of vehicles.
T_q = the duration of queue q.

Based on this, the following measures are computed.

- The maximum queue length over all queues given by:

$$L^{\max} = \underset{q}{\text{Max}}\ (L_q) \qquad\qquad\qquad\text{(A.66)}$$

- The mean of the maximum lengths of each queue over all queues given by:

$$\bar{L} = \frac{1}{Q}\sum_{q=1}^{Q} L_q \qquad\qquad\qquad\text{(A.67)}$$

- The standard deviation of the maximum queue lengths across all queues given by:

$$s_q = \sqrt{\frac{\sum\limits_{q=1}^{Q}(L_q - \bar{L})}{Q-1}}$$

(A.68)

- The maximum duration over all queues given by:

$$T^{\text{max}} = \text{Max}_q \, (T_q)$$

(A.69)

- The mean of the durations of each queue over all queues given by:

$$\bar{T} = \frac{1}{Q}\sum\limits_{q=1}^{Q} T_q$$

(A.70)

- The standard deviation of the durations across all queues given by:

$$s_T = \sqrt{\frac{\sum\limits_{q=1}^{Q}(T_q - \bar{T})}{Q-1}}$$

(A.71)

- The correlation between the maximum queue lengths and durations of queues given by:

$$\rho_{LT} = \frac{\sum\limits_{q=1}^{Q}(L_q - \bar{L})(T_q - \bar{T})}{\sqrt{\sum\limits_{q=1}^{Q}(L_q - \bar{L})^2}\sqrt{\sum\limits_{q=1}^{Q}(T_q - \bar{T})^2}}$$

(A.72)

These measures are computed independently for each queue type.

Time spent by vehicles in the queues

The *elemental observations* are the number of hours spent by each vehicle in a particular queue.

Let Q = the total number of queues of a particular type.

N = the total number of vehicles in a particular queue type.

123

h_{nq} = the number of hours spent in the queue q by vehicle n .

The measures are computed as follows.

- Total vehicle hours spent in the queues of a particular type by all vehicles given by:

$$H = \sum_{n=1}^{N} \sum_{q=1}^{Q} h_{nq}$$ (A.73)

These measures are computed independently for each type of queue and for vehicles of different status.

A.7 Frequency of Signal State Changes

The *elemental observations* for the performance measures are the number of signal state changes per minute measured over 1 minute intervals for each signal.

The measures capture the average and standard deviation of state changes for specific vehicles that exhibit the maximum and median mean number of state changes per minute. In addition, they capture the average and standard deviation of state changes over all signals.

Let M = the number of intervals (m = 1 to M).

      W = the total number of signals (w = 1 to W).

      F_{mw} = the number of signal state changes per minute of signal w for interval m.

Then, the mean number of signal state changes per minute for signal w is given by:

$$\overline{F}_w = \frac{1}{M} \sum_{m=1}^{M} F_{mw}$$ (A.74)

The maximum mean number of state changes is given by:

$$\overline{F}_{w^{max}}^{max} = \underset{w}{\text{Max}}\ (\overline{F}_w)$$ (A.75)

where w^{max} identifies the signal with the maximum mean.

The median mean number of state changes is given by:

$$\overline{F}_{w^{median}}^{median} = \underset{w}{\text{Median}} \, (\overline{F}_w)$$

(A.76)

where w^{median} identifies the segment with the median of mean.

Then, the measures are computed as follows:

• Mean number of state changes per minute for signal w^{max} as detailed by A.74.

• Standard deviation of the number of state changes per minute for signal w^{max} given by:

$$s_{w^{max}} = \sqrt{\frac{\sum_{m=1}^{M}(F_{mw^{max}} - \overline{F}_{w^{max}})^2}{M - 1}}$$

(A.76)

The above measures are similarly computed for signal w^{median}.

• Mean number of state changes per minute for all signals of a given type given by:

$$\overline{F} = \frac{1}{W}\sum_{w=1}^{W}\overline{F}_w$$

(A.77)

• Standard deviation of number of state changes per minute for all signals of a given type given by:

$$s_{\overline{F}} = \sqrt{\frac{\sum_{w=1}^{W}(\overline{F}_w - \overline{F})^2}{W - 1}}$$

(A.78)

These measures are computed for each signal type independently.

125

A.8 Vehicle Miles Traveled

The *elemental observations* are the distances traveled by each vehicle over its journey.

Let VMT_n = the total vehicle miles traveled by vehicle n.

   N = the total number of vehicles.

The measure of interest is obtained as follows:

• Total *VMT* across all the vehicles in the network given by:

$$VMT = \sum_{n=1}^{N} VMT_n \qquad\qquad (A.79)$$

A.9 Vehicle Hours Traveled

The *elemental observations* are the times spent by each vehicle over its journey.

Let VHT_n = the total vehicle hours traveled by vehicle n.

   N = the total number of vehicles.

The measure of interest is obtained as follows:

• Total *VHT* across all the vehicles in the network, given by:

$$VHT = \sum_{n=1}^{N} VHT_n \qquad\qquad (A.80)$$

Appendix B Methods for Solving Multiobjective Problems

Most of the classical problem-solving methods assume that compensatory tradeoffs can be established between objectives. Some very good examples of these methods are cost-effective analysis, benefit/cost analysis, and multiattribute utility theory [Rietveld 80, Roy 81].

The cost-effectiveness approach is concerned with how each alternative contributes to the attainment of the objectives. The performance measures serve as the dimension against which the alternatives are evaluated. The "effectiveness" of an alternative is usually represented as a scaled quantity relating to specific objectives. For example, some criteria for an alternative to be considered cost-effective may be that it cost no more than the base case alternative and give atleast 20% higher benefits on a specific objective. The disadvantage of this method is that it does not provide information on the relative value of alternatives. The best known approach to overcome this is an economic evaluation technique called the benefit/cost analysis.

In benefit/cost (B/C) analysis, the benefits and costs of alternatives are expressed in monetary terms and used to determine which alternative provided the most net benefit. An alternative with a B/C ratio of less than 1 is a likely candidate for rejection. A good alternative should have a B/C ratio of atleast 1. Although this is extensively used, there are problems associated with this method. In many cases, the definitions of costs and benefits are ambiguous and hence, the B/C ratio does not give good enough information for decision making. Also, conversion of benefits to monetary terms is not straight forward.

The multiattribute assessment method is based on expected utility theory. At the center of the method is the multiattribute utility function (which is obtained from revealed preferences) that assigns a unique value to any combination of levels for the various dimensions of the impacts (performance measures). The utility for a specific alternative is thus a function of the utilities of the various attribute levels associated with the alternative. Though this method is the most widely accepted formal methodology used in decision making in uncertain environment, it presents significant computational and conceptual problems. The main problem is the specification of the utility function which greatly influences the results of the assessment.

If the use of any of the above methods is justifiable, a problem can be solved as a single objective problem. The problem at hand clearly involves multiple conflicting performance categories, e.g. improving level of service, pollution minimization and improving safety conditions. Hence, it is extremely difficult to establish tradeoffs between them. Also, preferences among the categories are difficult to be specified because all of them are equally important. In addition, different people have different perceptions as to the relative importance of these categories. Because of these factors, the categories cannot be combined (i.e., reduced to a single objective), and consequently no single optimum solution is possible. Hence, a multiobjective decision making approach is chosen for the problem.

B.1 A Multicriteria Method for Selection or Refinement of Control Systems

B.1.1 The Multiobjective Decision Problem

A multiobjective decision problem is defined as a problem in which there is more than one objective and the objectives cannot be combined in any way [Giuliano 85]. Mathematically, it can be expressed as a vector optimization problem:

$$\max \; Z(x_1, x_2, .., x_n) = [Z_1(x_1, x_2, .., x_n), Z_2(x_1, x_2, .., x_n), ..., Z_p(x_1, x_2, ..., x_n)]$$
$$\text{subject to } g_i(x_1, x_2, ..., x_n) \leq 0 \quad \text{for} \quad i = 1, 2, ..., m \qquad (B.1)$$

where Z is the vector of objective functions $Z_1, Z_2, ..., Z_p$, g_i are the constraints and x_j are the decision variables. No single optimum solution is possible because either the objectives conflict or compensatory tradeoffs cannot be established between objectives. The objectives conflict in the sense that an optimal value of one objective implies non-optimal values for the other objectives. With respect to the problem at hand, it is not possible to find a control system that performs the best on all the performance categories.

Multiobjective methods of analysis seek to find the best possible alternatives. These are the best in that all others achieve the stated objectives to a lesser extent than those in the best possible set. Within the set, however, no single alternative can be judged the best, unless preferences between the objectives are specified. These best possible solutions are

128

called *non-inferior* solutions. Non-inferiority is analogous to the concept of Pareto optimality : A solution $Z(x)$ is non-inferior if there exists no feasible solution $Z(x')$ such that

$Z(x') \geq Z(x)$ and $Z_i(x') \neq Z_i(x)$ for atleast one *i*.

In most multiobjective problems, there are many non-inferior solutions. Identifying these is not sufficient since one solution must eventually be chosen. Therefore, the next step is to identify a smaller set of *compromise solutions* from which is one is chosen. Compromise solutions are those which are located closest to the ideal. The search for compromise solutions is based on the idea that "balanced" solutions are better than those which heavily favor one objective at the expense of other objectives. These solutions are favored because all the objectives come (relatively) close to reaching their maximums. At the end, a final choice is made from among the compromise alternatives.

As mentioned before, the problem at hand is a discrete multiobjective problem: A given set of alternatives is evaluated with respect to a set of objectives. Solving a discrete multiobjective problem entails choosing the best subset of alternatives from the finite set of feasible solutions. Most of the methods developed for these problems are based on a pairwise comparison of alternatives, which eventually leads to some sort of rank ordering. A more appropriate approach to the control system performance evaluation problem is to develop a method for identifying compromise solutions, solutions which are acceptable for different traffic conditions. The multiobjective evaluation method known as *concordance analysis* is adapted for the purpose.

B.1.2 Description of Concordance Analysis

Concordance analysis is a rank-ordering technique in which alternatives are ranked by a series of pairwise comparisons across the set of objectives. The original concordance technique, known as the Electre method, was developed in France [Giuliano 85]. The analysis is based on the project effects matrix, which contains a vector of scores for each alternative on each of the chosen objective measures. Two different indices are calculated from the project effects matrix. A concordance index calculates the degree to which one alternative is preferred to another for a given weighting structure on the objectives. A discordance index calculates the degree to which one alternative is dominated by another. Dominance indices are developed from the concordance and discordance scores, and they are used to establish the relative preference of each alternative with respect to the given

weighting scheme. Alternatives which perform better than average on both concordance and discordance are defined as non-dominated. The theoretical basis of concordance analysis is discussed in [Rietveld 80] and [Roy 81]. An application to transportation planning can be found in [Giuliano 85].

This method is extremely sensitive to the particular measures and weights chosen, because relative performance of the alternatives is a key factor in the analysis. Given a set of objectives, there may be a question about how many measures are appropriate. In particular, measures that are highly correlated with one another may be candidates for elimination, since it might be expected that little information would be lost in eliminating them. Therefore, the objectives and measures for inclusion in the analysis should be chosen with extreme caution. Also, it is preferable to have a unique measure for each objective, for ease of analysis. If more than one measure is perceived to be necessary, the weight given to the objective is distributed among the measures related to it.

B.1.3 Modification of Concordance Analysis

The two major inputs to the simulation are the controls system and the scenario. The different control systems are assumed to be the alternatives that need to be evaluated. Since the measures vary not only with the alternatives but also with the scenarios, both the measure and scenario dimensions should be used in determining the best compromise solutions. That is, the compromise solutions should perform "uniformly well" not only across the measures but also across the scenarios. However, in the event where scenarios pose severe constraints in identifying a compromise solution, different control systems can be used for different sets of scenarios as the "best" alternatives.

Appendix C   Control strategies for Objectives

The control strategies presented in this appendix represent a first cut at the control strategy development effort. Hence, they need to be debugged and refined before implementation in case study.

C.1 Control Strategy for 'Reducing Travel Time'

Goal: Improve Level of Service
Control Function: Access Control

For this objective, access to the mainline is controlled using on-ramps. The principle of entrance ramp control is the limiting of the number of vehicles entering the mainline so that the demand on the mainline itself will not exceed its capacity. The state-of-the-art in ramp metering in design and practice are known to differ. The types of entrance ramp control that are in common use are:
- Closure
- Pretimed ramp metering
- Traffic-responsive ramp metering
- Gap acceptance merge control
- Integrated ramp control.

Traffic responsive ramp metering is chosen for implementation due to its capability of accounting for (adapting to) real-time changes in traffic conditions. Traffic-responsive metering is directly influenced by the mainline and ramp traffic conditions during the metering period. Metering rates are selected based on the real-time measurements of traffic variables indicating the current relation between upstream demand and downstream capacity.

Variations on the basic strategy of traffic-responsive metering utilize different combinations of traffic variables. The principle traffic-responsive strategies are as follows:
- Demand-Capacity control

131

- Occupancy control.

Demand-capacity control type of traffic responsive ramp metering is used for reducing travel time. Also, the metering rates for each ramp are obtained individually signifying the local nature of the control.

Demand-Capacity Control

The exact travel times of vehicles between any two points are not available until the completion of the simulation. However, during the simulation, an increase in the travel time on a link has to be known for the activation of the control strategy for 'Reducing travel time'. The decision as to when to activate the control strategy is made using estimated travel times.

In each time interval k, for each of the segments on the simulated network, the travel time obtained is used as the estimate of travel time on the segment for the next time interval. Let this be denoted by t_s. If t_s exceeds the free-flow travel time for the segment, it is concluded that the travel time on the segment has increased and hence, the control strategy to reduce the travel time needs to be activated.

Metering rates are determined based on a real-time comparison of upstream volume and downstream capacity. Let P denote the point at which the on-ramp intersects the mainline. The flow rate at the detector station on the mainline immediately upstream of P gives the volume (vehicles per hour) upstream of the ramp (U). The downstream capacity (C) can be obtained as follows :

The capacity is affected by weather conditions, traffic conditions and incidents. The mainline section considered for simulation has only two lanes in almost all of its length. And the scenarios considered assume normal weather conditions. Hence, the capacity needs to be calculated under three different conditions:

- No lane blockage : If there is no lane blockage, then the flow rate at the end of the segment (the last loop detector on the segment) immediately downstream of P is used as the downstream capacity.
- One lane blockage : For a one lane blockage, the effective capacity is assumed to be 1630 passenger cars per hour.
- Two lane blockage : If both the lanes are blocked, then the vehicles cannot pass through the incident location and hence, the capacity is reduced to zero.

132

The difference between the upstream volume and downstream capacity is used as the allowable entrance ramp volume, denoted by X. Therefore,

$$X = C - U.$$

The ramp volume X is expressed as a metering rate to be used in the next control interval (one minute). Depending upon this metering rate, single-entry metering or platoon metering is used. In the case of single-entry metering, the ramp metering signal is timed to permit only one vehicle per green interval. Platoon metering permits 2 or more vehicles to be released per cycle.

The metering rates are decided as follows:

- If $X \geq 1100$ veh/hr, the ramp vehicles are allowed free access to the mainline, that is, no metering is used at the ramp.

- If $300 \leq X \leq 1100$ veh/hr, appropriate metering rates are used. The number of cycles per minute and the green, yellow and red times in each cycle are determined based on the entries in Table C-1. A green-plus-yellow (or just green if yellow is not used) time of 3 seconds is assumed to be just long enough for one vehicle to proceed past the signal.

- If $X \leq 300$ veh/hr, a minimum metering rate of 4 veh/min[1] is used. 4 cycles per minute with a cycle time of 15 seconds and a green, amber and red times of 3, 0, 10 seconds are used respectively.

- If $C = 0$, then the ramp is closed for the time interval using a red signal.

---

[1] A metering rate lower than 4 veh/min is found ineffective [TCSH 85], because vehicles waiting at the ramp will judge the metering signal to be malfunctioning and proceed through on red.

Table C-1 Ramp signal timings when metering rates lie between 300 and 1100 vph

Vph	Vpm	No. of cycles	Cycle Time	Green Time	Yellow Time	Red Time
300-330	5	5	12	3	0	9
330-390	6	6	10	3	0	7
390-450	7	7	8.57	3	0	5.37
450-510	8	4	15	4	2	9
510-570	9	3	20	7	2	11
570-630	10	5	12	4	2	6
630-810	11-13	6	10	4	2	4
810-870	14	7	8.57	4	2	2.57
870-930	15	3	20	13	2	5
930-990	16	4	15	10	2	3
990-1100	17-18	3	20	16	2	2

Pseudo code

In each time interval k,

    For each segment,

        Obtain the travel time on the segment, t_s.

        Let t_f be the free-flow travel time for that segment.

        If $(t_s > t_f)$,

            Conclude that the travel time has increased.

            Let U denote the volume (vph) upstream of the ramp

                C denote the downstream capacity (vph)

                X denote the allowable entrance ramp volume.

        Then, X = C - U.

        If   X ≥ 1100 veh/hr, the ramp vehicles are allowed free access to the mainline

                : no metering.

        300 ≤ X ≤ 1100 veh/hr, appropriate metering rates as in the table are used.

        X ≤ 300 veh/hr, a minimum metering rate of 4 veh/min is used.

        C = 0, the ramp is closed.

            This rate is used for the next control interval.

134

C.2 Control Strategy for 'Reducing Time Spent in Queues'

Goal: Improve Level of Service

Control Function: Mainline Control

While designing the control strategy for reducing the time spent in queues, the following should be taken into consideration:

- whether the queue is formed with lane blockage or without lane blockage.
- whether it is a single lane blockage or a double lane blockage (the network used for the case study has single and double lane segments all along except at the merging and diverging points).

Queue without lane blockage: any queue (recurrent congestion)

A stretch of roadway (freeway or tunnel) is divided into sections based on the Variable Speed Limit Signs (VSLS). The zone of influence of each VSLS is considered to be a section.
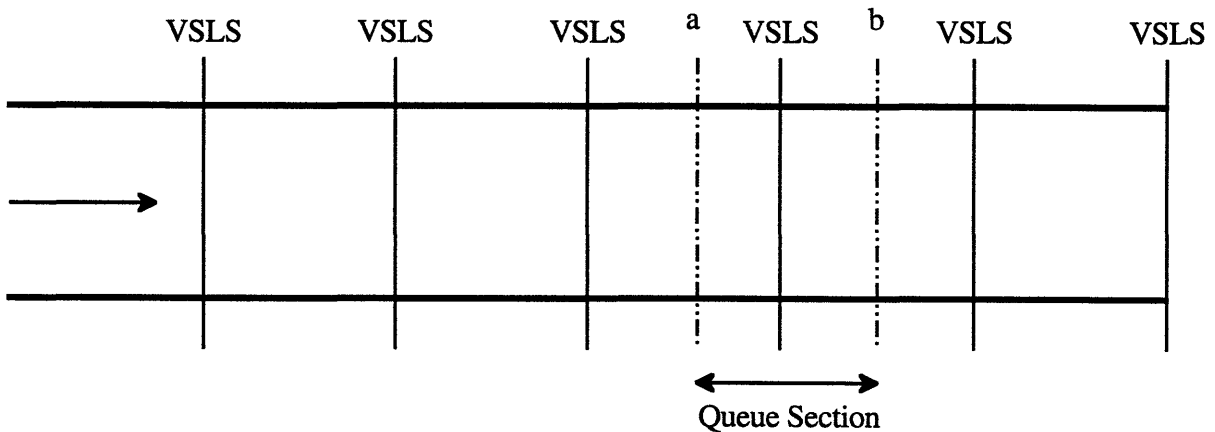

Figure C-1 A roadway section

In Figure C-1, section ab, where b is the beginning of the queue and a is the end of the queue, is the bottleneck section where the queue is assumed to have formed. The objective is to minimize the time spent in the queue by controlling the traffic flow using VSLS and possibly, MM.

135

The strategy involves the creation of an increasing speed gradient along the flow direction upstream of a. This postpones the arrival of vehicles at a. The speeds should be fixed such that by the time vehicles reach a, the queue is on the verge of complete dissipation. If the dissipation rate of the queue is known, the time for complete clearance of the queue can be obtained. Based on this, the speeds upstream can be adjusted to ensure that the vehicles will reach a only when the queue is nearing complete clearance. The approach used is as follows:

Figure C-2 is used as reference to the description of the approach.
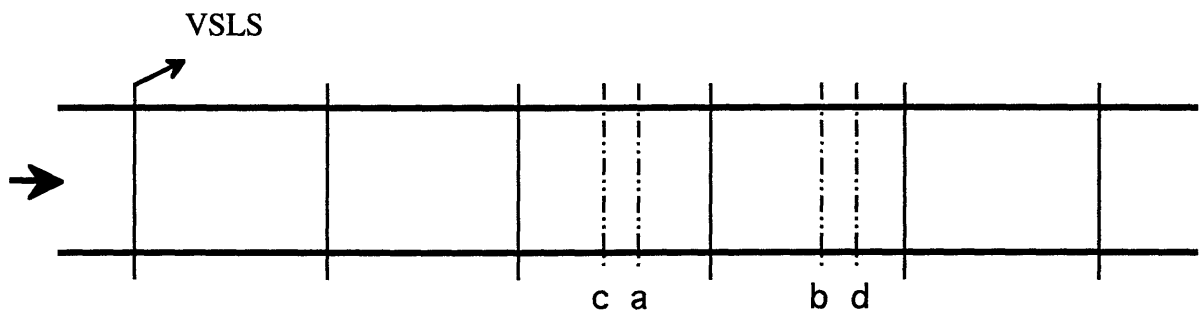
VSLS

Figure C-2   Reference figure for queue without lane blockage

As soon as a queue is detected, the beginning (b) and the end (a) of the queue are determined. The flow of vehicles can be measured only at the detector locations. Let c denote the position of the detectors immediately upstream of a and d denote the detector location immediately downstream of b. Let the flow rates (vehicles per hour) at c and d be denoted them by F_c and F_d. F_c is greater than F_d since there is a queue formation on ab.

The first step is to reduce the flow rate at c to less than that at d, so that the queue will slowly get cleared up. This is done by reducing the speeds in the upstream sections. The flow rate at c is controlled to be a *fraction (less than 1)* of the flow rate observed at d. Let this required flow rate be F'_c. The speed required to maintain this flow is obtained by substituting F'_c in the speed-flow relationship for the link. The vehicles just upstream of a will be traveling very slowly due to the presence of the queue downstream. Hence, the speed limit sign which is just upstream of a (within τ ft of a) cannot be used to effectively control the speeds of vehicles approaching the queue. Therefore, if there is a VSLS within τ ft of a, then it should indicate a stopping speed of 15 mph (since the vehicles near the

queue are anyway traveling very slowly). The VSLS immediately upstream of this (say, VSLS1) is set to the speed limit obtained by substituting the desired flow rate in the speed-flow relationship. Since the speed limits are multiples of 5, the speed obtained is rounded off to the lower multiple of 5 (say S), to be on the safer side. If there is no VSLS within τ ft of a, then the VSLS immediately upstream of a (VSLS1) is set to the rounded off speed (S). The speed on the VSLS immediately upstream of VSLS1 (VSLS2) is set to (S-S_{diff}) or 15 mph, whichever is greater. Once the queue is cleared up, the flow rates are made equal by imposing same speeds on all sections.

Let the average speed of vehicles on the section upstream of VSLS2 be S_{us}. This can be obtained as the average of the instantaneous speeds of all vehicles on the section in the middle of the previous time interval. It is important to ensure that the speed difference between two subsequent sections is no more than 10 mph, to prevent drastic acceleration/deceleration by drivers. Two cases are possible: the speed upstream is much greater than VSLS2 or VSLS2 is much greater than speed upstream. In such a case, the speed limits are adjusted as follows:

If (S_{us} - VSLS2) > 10

    $VSLS2_{new}$ = S_{us} - 10, rounded off to the next lower multiple of 5.

    If ($VSLS2_{new}$ - VSLS1) > 10

        $VSLS1_{new}$ = $VSLS2_{new}$ - 10

If (VSLS2 - S_{us}) > 10

    $VSLS2_{new}$ = S_{us} + 10, rounded off to the next lower multiple of 5.

    If (VSLS1 - $VSLS2_{new}$) > 10

        $VSLS1_{new}$ = $VSLS2_{new}$ + 10

A queue is formed because the demand is higher than capacity. Hence, the reduction of speeds upstream of the queue reduces the distance between the vehicles and under congested conditions, may lead to the formation of a new queue. If a new queue is detected, the length is found out. The queue length of this queue is constantly updated and if it exceeds a pre specified number of vehicles (Q), the traffic flow is regulated upstream of this queue. The first occurring Mainline Meters (LUSs) at least X ft upstream of the end of the new queue are used to block the traffic. As soon as the queue length exceeds Q, the mainline meters are turned yellow and after a time interval t, turned red. The time of reopening of mainline meters is decided as follows :

    Let T_q denote the time for the original queue to get completely cleared.

    Let x denote the position of the mainline meters used.

137

Let T_{xb} denote the time taken by the first vehicle behind the mainline meters to reach the beginning of the original queue, b. The speed of this vehicle is assumed to be slightly higher than the speed of the vehicles clearing from the queue. The average speed at d is multiplied by a *Speed Factor* (greater than 1). This speed factor should incorporate in it the startup delay of the first vehicle and the empty space upstream of this vehicle which will allow it to travel faster than the queued vehicle.

Let T_{cur} denote the current time.

Then, the mainline meters are reopened at $T_{cur} + T_q - T_{xb}$.

T_q can be obtained as follows:

Let F_d denote the flow rate at d (vehicles per hour) measured for the time interval before turning the mainline meters red.

Let N_{xb} denote the number of vehicles between x and b, immediately after the mainline meters are closed.

Then, $T_q = N_{xb}/F_b$.

The pseudo code for the above strategy is given below.

Pseudo code for control strategy for 'No lane blockage / any queue'

As soon as a queue is detected,

Obtain the beginning (b) and end (a) of the queue.

Let c denote the detector location immediately upstream of a and d, the detector location immediately downstream of b.

Obtain the flow rates at c and d. (F_c and F_d)

Make $F_c = KF_d$, where $0 \leq K \leq 1$.

Substitute F_c in the speed-flow relationship for that link and obtain the speed.

Round off this speed to the next lower multiple of 5. (say, S)

If there is a VSLS within τ ft upstream of a ($VSLS_a$),

then the speed limit sign immediately upstream of $VSLS_a$ = VSLS1.

Use a stopping speed of 15 mph on $VSLS_a$.

If there is no VSLS within τ ft upstream of a,

then the speed limit sign immediately upstream of a = VSLS1.

Use the rounded off speed limit as the limit on the zone influenced by VSLS1.

Change speed limit on the section immediately upstream of VSLS1 (VSLS2) to $S-S_{diff}$.

138

Let the average speed of vehicles on the section upstream of VSLS2 be S_{us}.

    If $(S_{us} - VSLS2) > 10$

        $VSLS2_{new} = S_{us} - 10$, rounded off to the next lower multiple of 5.

        If $(VSLS2_{new} - VSLS1) > 10$

            $VSLS1_{new} = VSLS2_{new} - 10$

    If $(VSLS2 - S_{us}) > 10$

        $VSLS2_{new} = S_{us} + 10$, rounded off to the next lower multiple of 5.

        If $(VSLS1 - VSLS2_{new}) > 10$

            $VSLS1_{new} = VSLS2_{new} + 10$

If a new queue forms upstream of a,

    If this queue length exceeds a fixed number Q,

        Turn the first mainline meter that is at least X ft upstream of the end of the new queue (x) to yellow.

        After a time period t, change the mainline meters from yellow to red.

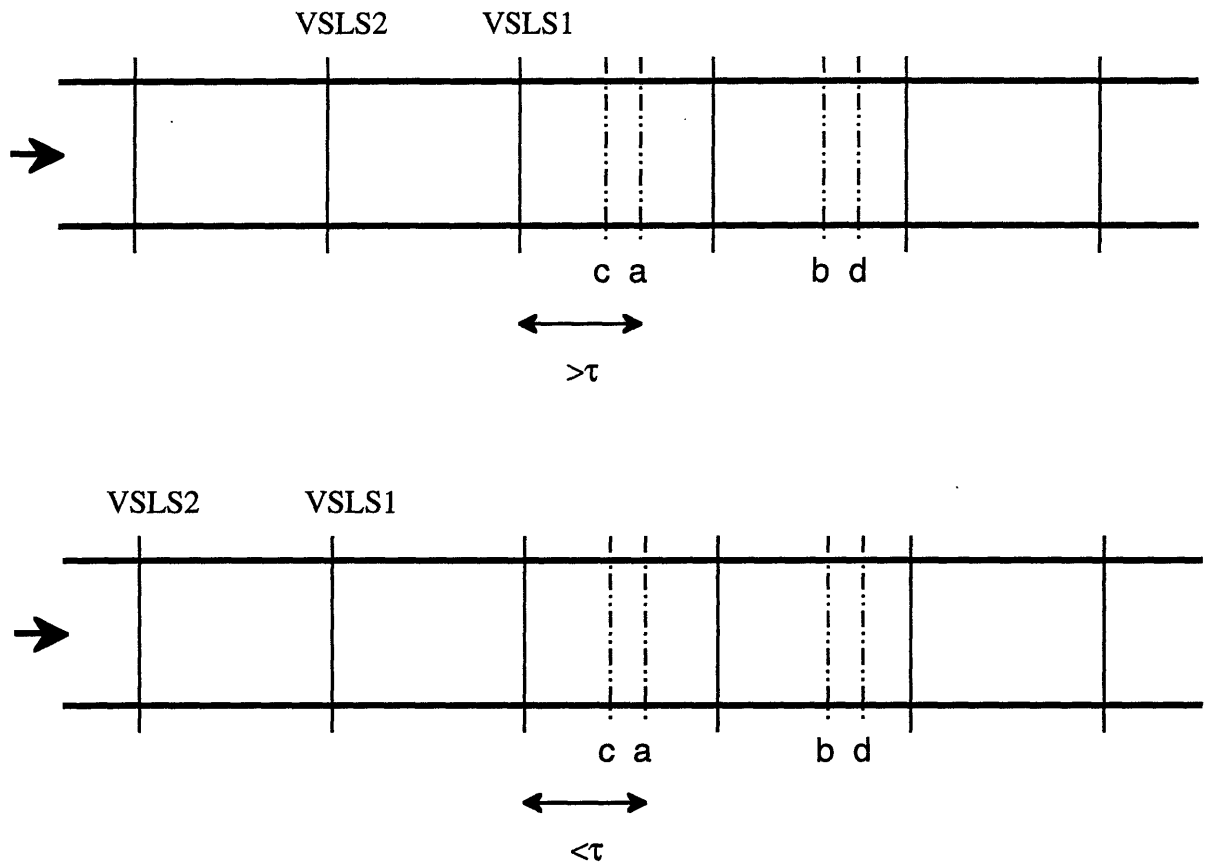   The mainline meters are reopened at $T_{cur} + T_q - T_{xb}$

VSLS2   VSLS1

c a   b d

>τ

VSLS2   VSLS1

c a   b d

<τ

Figure C-3 VSLS section on which the obtained speed is to be used

Queue with lane blockage (non-recurrent congestion)

The roadway section used for the demonstration has two lanes in each direction. So, the control system designed is for a roadway with two lanes. The discussion is divided into two sections - regarding left lane blockage and right lane blockage.

Left lane blockage: All the ramps to the mainline in the network used for the experiment are adjoining the right lane. Hence, the effect of ramps on the traffic due to left lane blockage is considered negligible.

When a lane blockage occurs, the speeds of vehicles near the incident automatically decreases since the capacity reduces drastically. Beginning at the incident location, the queue starts building up. The objective is to clear the vehicles from the

queue as soon as possible and also prevent the queue buildup once it is cleared - that is ensure a smooth passage of arriving vehicles (at the incident location) until incident is cleared.

The control strategy has two phases. Initially, an 'upstream yellow length' (UYL), 'upstream red length' (URL), and 'downstream red length' (DRL) are identified. They represent the sections on which the Lane Use Signs (LUS) will be turned to yellow or red. In the first phase, the LUSs on all the identified lengths are changed to yellow giving a warning to the traffic to be wary and be prepared to change lanes. In the second phase, the LUSs on the upstream and downstream red lengths are changed to red and the LUSs on the upstream yellow length are retained on yellow. For this case of left lane blockage and all other cases considered subsequently, the following distances are used for each of the above lengths :

Upstream Yellow Length : 1000'
Upstream Red Length : 1000'
Downstream Red Length : 400'.

When dealing with a straight stretch of road without the presence of on/off ramps near the incident location, by default, a stopping speed of 15 mph is used as speed limit on sections with Red LUS and a merging speed of 25 mph is used a speed limit on sections with Yellow LUS. To incorporate the effect of on/off ramps, these speed limits have to altered slightly. The speed limits and LUSs to be set on the respective sections in phase 1 and phase 2 for a left lane blockage are shown in Figures C-4 and C-5 respectively.
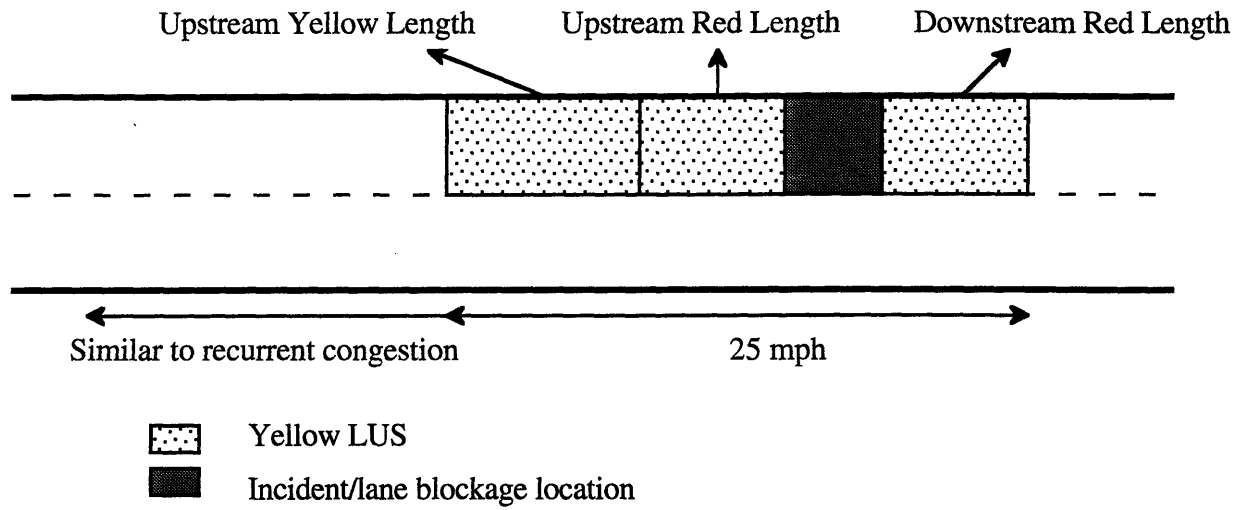
Upstream Yellow Length   Upstream Red Length   Downstream Red Length

Similar to recurrent congestion   25 mph

Yellow LUS

Incident/lane blockage location

Figure C-4   Left lane blockage : speeds in phase 1

Upstream Yellow Length   Upstream Red Length   Downstream Red Length

Similar to recurrent congestion   25 mph   15 mph

Yellow LUS

Upstream/Downstream Red LUS
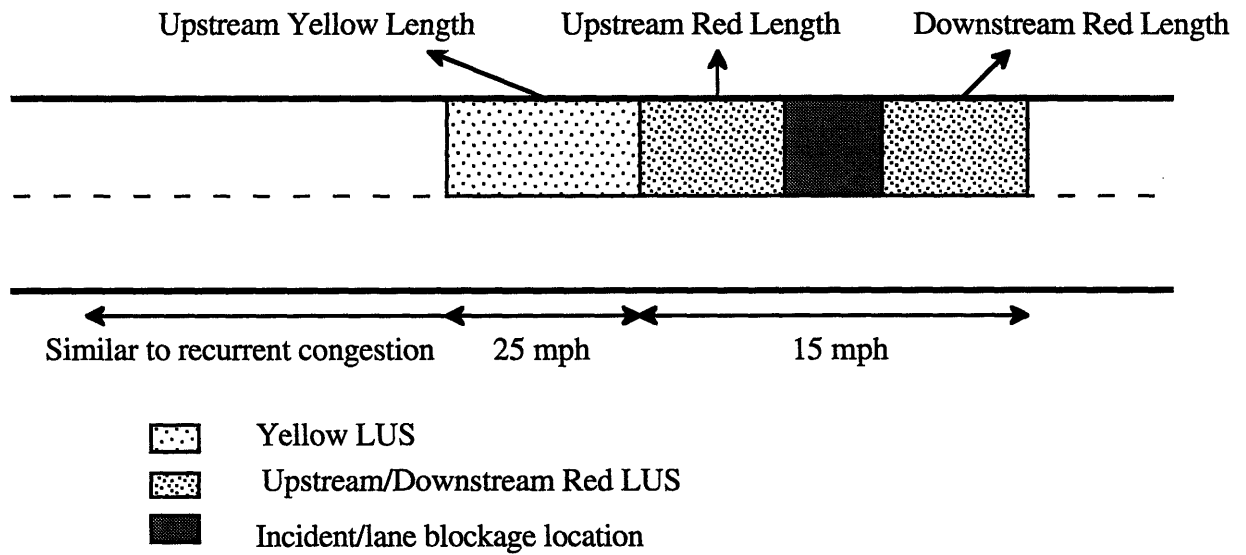
Incident/lane blockage location

Figure C-5   Left lane blockage: speeds in phase 2

142

Figure C-6 is used as reference to the description of the control strategy below.
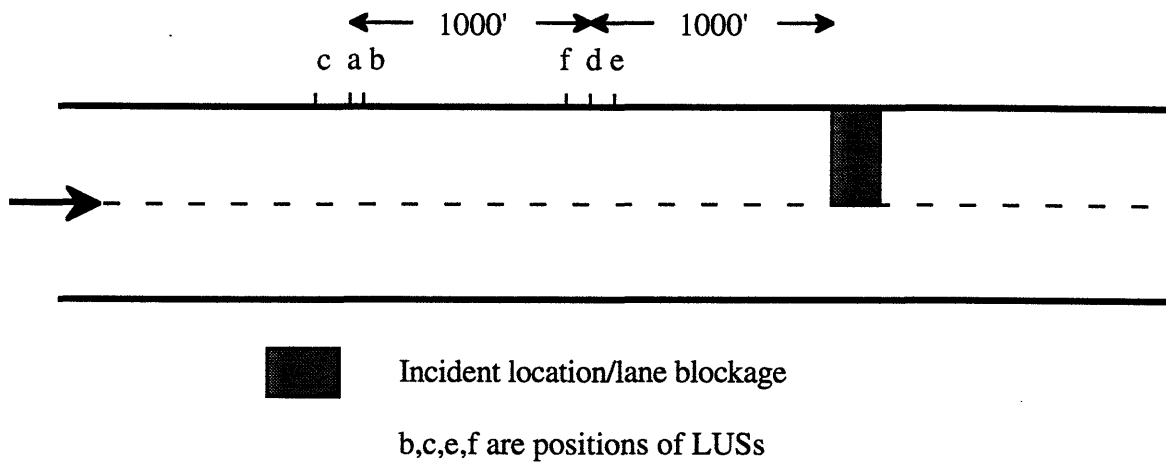

Figure C-6 Reference figure for left lane blockage

In the first phase, all the LUSs in the 2000' upstream of the incident location (sum of the upstream yellow and red lengths) are turned yellow. Let a represent the point 2000' upstream of the incident location. Let b represent the position of the LUS nearest to a within the 2000'. There is no LUS regulating the traffic between a and b. If the distance ab is less than $X1$ ft, then b is taken to be the last LUS in operation (which is yellow) due to the incident. If the distance ab is greater than $X1$ ft, then the LUS immediately upstream of b (indicated by c) is turned to yellow. In the DRL of 400', all the LUSs are turned to yellow.

If there is a VSLS within $X2$ ft of the incident location (let us represent this by $VSLS_{X2}$), then this VSLS cannot govern the speeds of the vehicles in the URL. Hence, the two VSLSs immediately upstream of $VSLS_{X2}$ are used to indicate the merging speed of 25 mph. $VSLS_{X2}$ will also indicate a speed limit of 25 mph to regulate the traffic on DRL. If there is no VSLS within $X2$ ft of the incident location, then the two VSLSs immediately upstream of the incident location will indicate the merging speed of 25 mph. Any VSLS in the DRL will also indicate a speed limit of 25 mph.

In the second phase, a similar method is followed. In the URL of 1000', all the LUSs are turned red. Let d represent the point 1000' upstream of the incident location. e represents the position of the LUS nearest to d within the URL and f represents the position of LUS immediately upstream of d. If de is less than $X1$ ft, then e will mark the

143

last LUS indicating red and f will mark the first LUS indicating yellow upstream of the incident location. If de is greater than $X1$ ft, then f will mark the last LUS indicating red and the LUS immediately upstream of f will mark the first LUS indicating yellow upstream of the incident location. The last LUS indicating yellow upstream of the incident location will be b or c as decided in the first phase. In the DRL of 400', all the LUSs are turned to red.

If there is a VSLS within $X2$ ft of the incident location (let us represent this by $VSLS_{X2}$), then this VSLS cannot govern the speeds of the vehicles in the URL. Hence, the VSLS immediately upstream of $VSLS_{X2}$ (let us represent this by $VSLS_{red}$, to indicate that it governs the speed on URL) is used to indicate the stopping speed of 15 mph and the VSLS immediately upstream of $VSLS_{red}$ is used to indicate the merging speed of 25 mph. $VSLS_{X2}$ will also indicate a speed limit of 15 mph to regulate the traffic on DRL. If there is no VSLS within $X2$ ft of the incident location, then the VSLS immediately upstream of the incident location will become $VSLS_{red}$ and will indicate the stopping speed of 15 mph and the VSLS immediately upstream of this $VSLS_{red}$ is used to indicate the merging speed of 25 mph. Any VSLS in the DRL will indicate a speed limit of 15 mph.

Let us denote the length of the first phase (time) by T. The second phase lasts until the incident is completely cleared when all the LUS are turned to green.

Upstream of UYL and downstream of DRL, the speed limits are decided based on the observed traffic conditions. Upstream of UYL, if there is a queue formation, the control strategy used will be similar to that for recurrent congestion.

The pseudo code for the above strategy is given below.

Pseudo code for control strategy for 'left lane blockage'

If a lane blockage is detected,
    Let L be the incident location.
    Identify the UYL, URL and DRL.
    First Phase
        a is a point 2000' upstream of L.
        b is the LUS farthest from L within the 2000' upstream.

c is the LUS immediately upstream of b.

If (ab < X1)

Turn all the LUSs upstream of L upto and including b to yellow.

If (ab > X1)

Turn all the LUSs upstream of L upto and including c to yellow.

Turn all LUSs in DRL to yellow.

Let S_1 represent the first VSLS upstream of L.

If ($S_1L < X2$)

Use a speed limit of 25 mph on S_1 and the two VSLSs immediately upstream
of S_1.

If ($S_1L > X2$)

Use a speed limit of 25 mph on the two VSLSs immediately upstream of L.

Use a speed limit of 25 mph on any VSLS in DRL.

The first phase lasts for time T.

Second Phase

d is a point 1000' upstream of L.

e is the LUS farthest from L with the 1000' upstream.

f is the LUS immediately upstream of e.

LUS_{last} is the farthest LUS upstream of L (b or c) used in the first phase.

If (de < X1)

Turn all the LUSs upstream of L upto and including e to red.

Turn all the LUSs upstream of e upto and including LUS_{last} to yellow.

If (de > X1)

Turn all the LUSs upstream of L upto and including f to red.

Turn all the LUSs upstream of f upto and including LUS_{last} to yellow.

Turn all LUSs in DRL to red.

If ($S_1L < X2$)

Let $VSLS_{red}$ be the VSLS immediately upstream of S_1.

Use a speed limit of 15 mph on $VSLS_{red}$.

Use a speed limit of 25 mph on VSLS immediately upstream of $VSLS_{red}$.

If ($S_1L > X2$)

Use a speed limit of 15 mph on S_1.

Use a speed limit of 25 mph immediately upstream of S_1.

Use a speed limit of 15 mph on any VSLS in DRL.

If there is a queue formation upstream of UYL,

Use the control strategy for recurrent congestion.

145

Right lane blockage: All the ramps to and from the mainline are adjoining the right lane. Depending upon whether the lane blockage is before/after on/off ramp, the effect on the traffic movement will be different. Hence, each of the cases is considered independently. A lane blockage is termed to be before/after on/off ramp if it occurs within X_{Ramp} ft of the point at which the ramp intersects the mainline.

Right lane blockage : before on-ramp

Let P denote the point at which the on-ramp intersects the mainline and L, the incident location. If (LP < X_{Ramp}), then the blockage is said to occur before on-ramp. Figure C-7 below illustrates a right lane blockage before on-ramp. A control strategy similar to the one used for the left lane blockage may be used for this scenario. The yellow and red lengths are obtained and the LUSs are implemented on the right lane instead of left lane. The speed limits imposed on the different lengths are the same as that done for a left lane blockage and are given as follows:

	Phase 1	Phase 2
Upstream Yellow Length	25	25
Upstream Red Length	25	15
Downstream Red Length	25	15

If P falls under the downstream red length, then the vehicles coming from the on-ramp should be given free access to the mainline and advised to tread carefully.
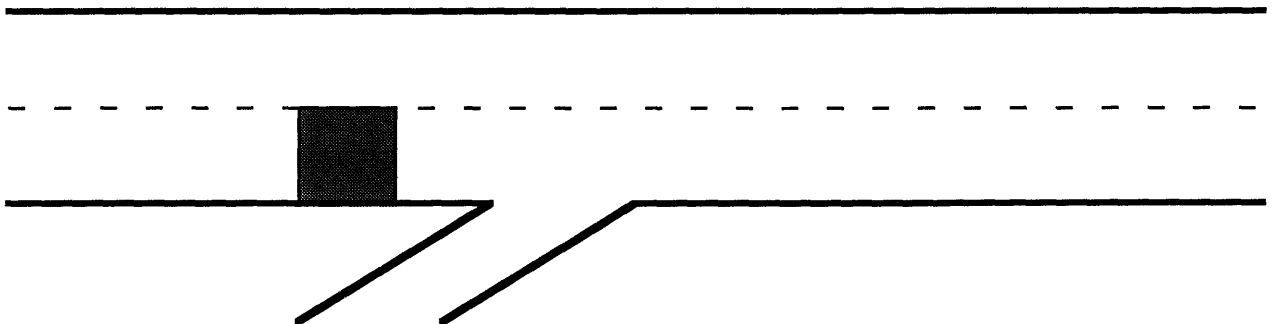

Figure C-7 Right lane blockage before on-ramp
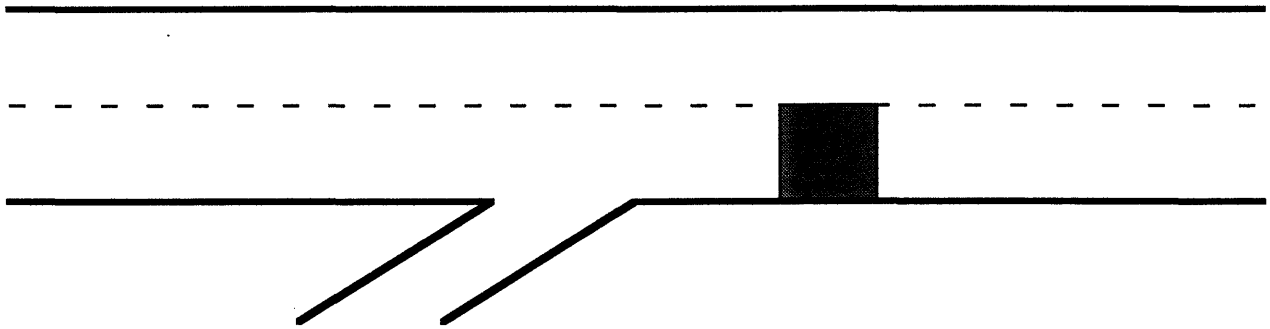
Right lane blockage : after on-ramp

Figure C-8 Right lane blockage after on-ramp

Let P denote the point at which the on-ramp intersects the mainline and L, the incident location. If (LP < X_{Ramp}), then the blockage is said to occur after on-ramp. Figure C-8 illustrates a right lane blockage after on-ramp. A right lane blockage after the on-ramp blocks the traffic entering the mainline from the on-ramp. The control strategy similar to the one used for left lane blockage may be used for this scenario. The yellow and red lengths are obtained and the LUSs are implemented on the right lane instead of left lane. The portion of the roadway around P is susceptible to incidents due to merging and lane changing. Therefore, the control strategy is used with the following changes to the VSLS:

- If P is part of the upstream red length,
    - in the first phase, the speed limits in both the upstream red and yellow lengths are reduced by an additional 5 mph to 20 mph.
- If P is part of the upstream yellow length,
    - in both the phases, the speed limit in the upstream yellow length is reduced by an additional 5 mph to 20 mph.

	P in URL		P in UYL	
	Phase 1	Phase 2	Phase 1	Phase2
Upstream Yellow Length	20	25	20	20
Upstream Red Length	20	15	25	15
Downstream Red Length	25	15	25	15

Right lane blockage : before off-ramp
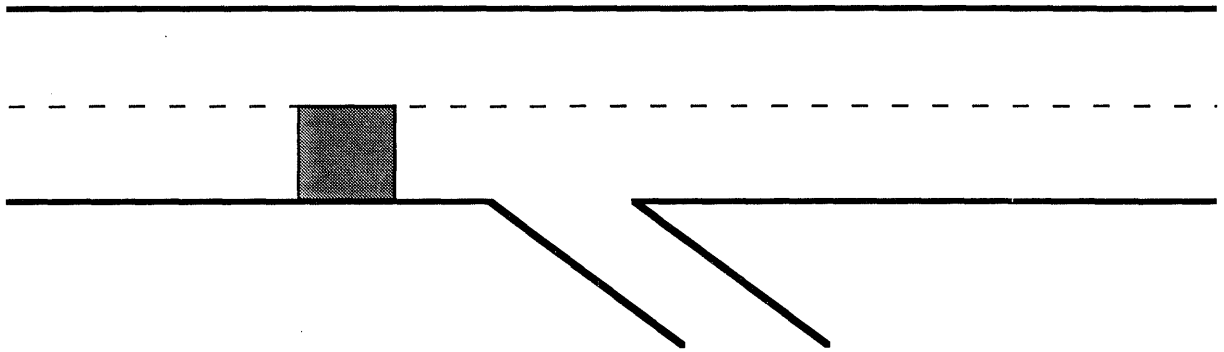

Figure C-9   Right lane blockage before off-ramp

Let P denote the point at which the on-ramp intersects the mainline and L, the incident location. If (LP < X_{Ramp}), then the blockage is said to occur before off-ramp. Figure C-9 illustrates a right lane blockage before off-ramp. The control strategy similar to the one used for left lane blockage may be used for this scenario. The yellow and red lengths are obtained and the LUSs are implemented on the right lane instead of left lane. If the lane blockage occurs immediately before the off-ramp, then P falls under the downstream red length. As a result, vehicles that need to use the off-ramp will be forced to miss the off-ramp. So, the speeds in this zone will have to be reduced additionally by 5 mph to 20 mph in the first phase and only the vehicles that need to use the off-ramp should be allowed into the downstream red section.

	P in DRL		P after DRL	
	Phase 1	Phase 2	Phase 1	Phase2
Upstream Yellow Length	25	25	25	25
Upstream Red Length	25	15	25	15
Downstream Red Length	20	15	25	15

Right lane blockage : after off-ramp

Let P denote the point at which the on-ramp intersects the mainline and L, the incident location. If (LP < X_{Ramp}), then the blockage is said to occur after off-ramp. Figure C-10 illustrates a right lane blockage after off-ramp. The control strategy similar to the one used for left lane blockage may be used for this scenario. The yellow and red lengths are obtained and the LUSs are implemented on the right lane instead of left lane.
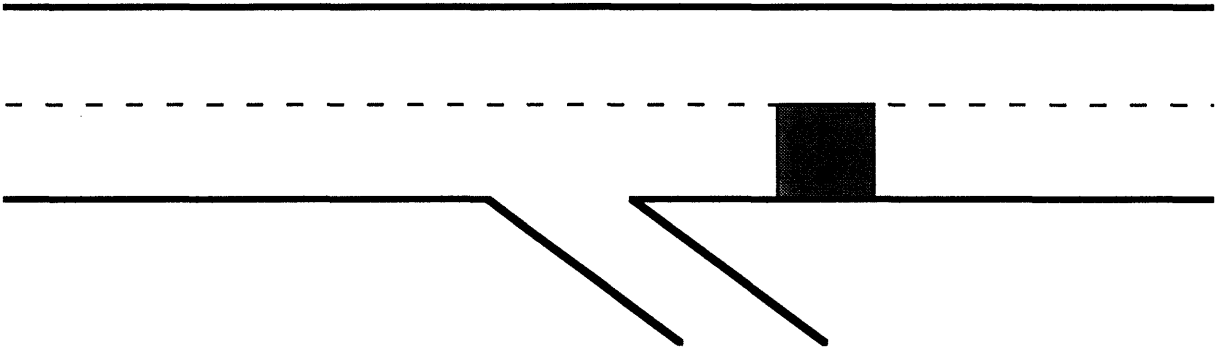

Figure C-10   Right lane blockage after off-ramp

The vehicles on the left lane that have to use the off-ramp have to make a right lane change and the vehicles on the right lane that have to go straight have to make a left lane change resulting in a high disturbance in this section. Hence, the control strategy is implemented with the following changes:

- If P is part of the upstream red length,
    - in the first phase, the speed limits in both the upstream red and yellow lengths are reduced by an additional 10 mph to 15 mph.
    - in the second phase, the speed limits in both the upstream red and yellow lengths are kept the same as in the first phase.
- If P is part of the upstream yellow length,
    - in the first phase, the speed limit in only the upstream yellow length is reduced by an additional 10 mph.
    - in the second phase, the speed limit in the upstream yellow length is kept the same as in the first phase.

	P in URL		P in UYL	
	Phase 1	Phase 2	Phase 1	Phase2
Upstream Yellow Length	15	15	15	15
Upstream Red Length	15	15	25	15
Downstream Red Length	25	15	25	15

Two lanes blocked

When both the lanes are blocked due to an incident, then, all the vehicles upstream of the incident will get stalled before the incident location. Hence, as soon as the incident is detected and confirmed to be a two lane blockage, the distances corresponding to the three lengths - UYL, URL and DRL are obtained and the LUS and VSLS in each of the lengths changed accordingly. The control strategy is similar to that of a left lane blockage except that the LUSs are now implemented on both the lanes. The speed limits in the different lengths in the two phases are as follows:

	Phase 1	Phase 2
Upstream Yellow Length	25	15
Upstream Red Length	15	15
Downstream Red Length	25	15

Figures C-11 and C-12 illustrate the LUS and VSLS changes in the two phases.
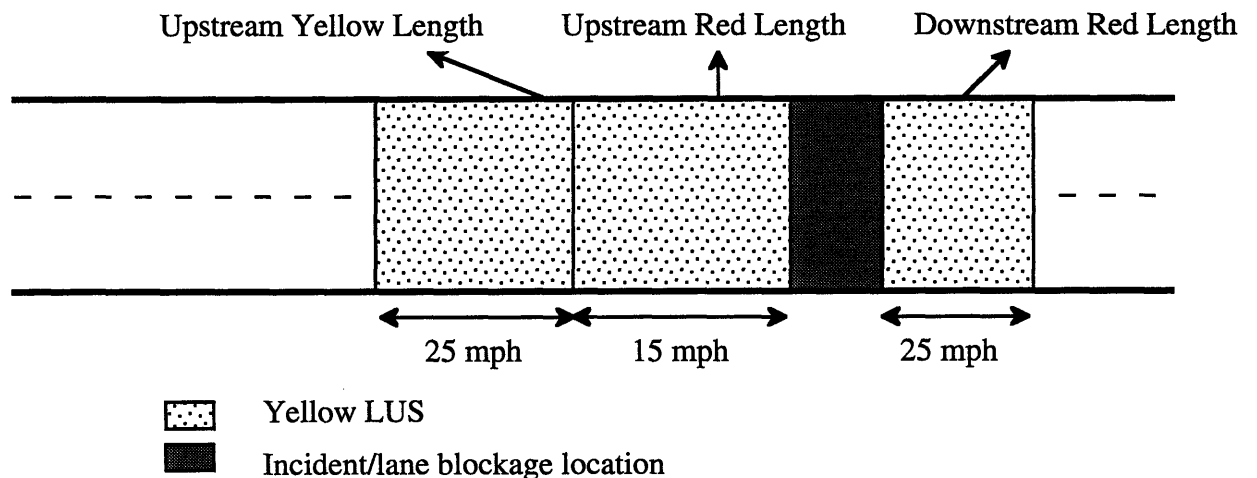

Figure C-11   LUS and VSLS changes in phase 1 for two lane blockage
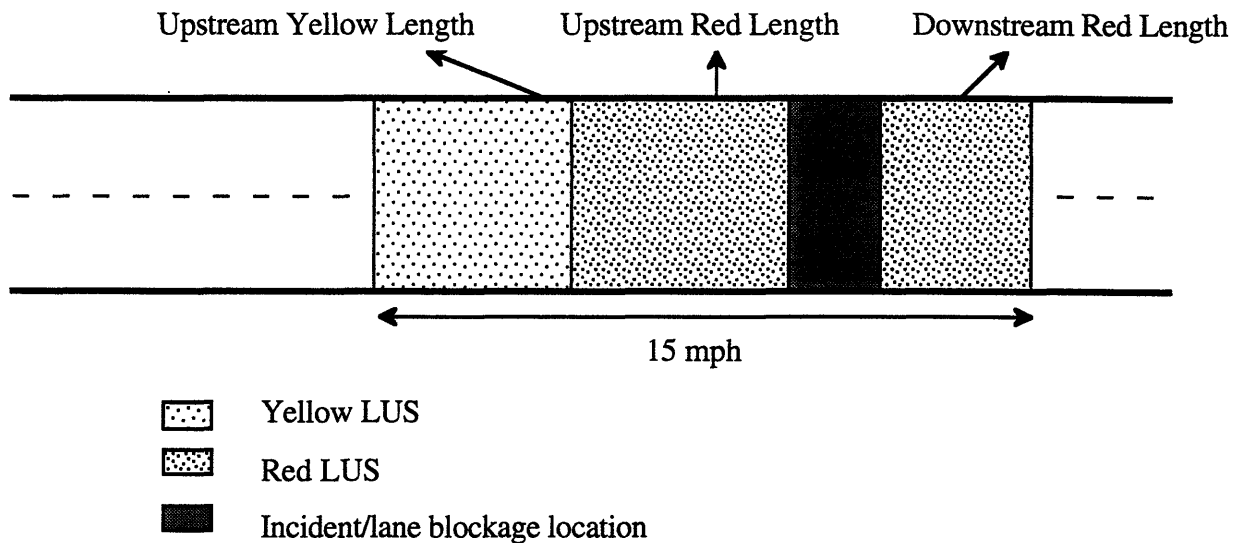
150

Figure C-12   LUS and VSLS changes in phase 2 for two lane blockage

C.3 Control Strategy for 'Reducing Link Density'

Goal: Reduce CO emissions and exposure
Control Function: Mainline Control

Most part of the network considered is an underground section (tunnel). The CO levels in the tunnel are mainly influenced by the number of vehicles and the acceleration and deceleration of the vehicles. Hence, one way of reducing the CO levels is to reduce the number of vehicles in the tunnel. The tunnel section used for demonstration has two lanes. Hence, the control strategy designed is for a tunnel with two lanes. Three different scenarios are possible:

- **No lane blockage** : In this case, no lanes in the tunnel are blocked and hence, the congestion and the resulting increased link density is due to the demand exceeding capacity.

- **One lane blockage** : One of the lanes in the tunnel is blocked in this case. The blocked lane can be either the left one or the right one. The incremental effect on the density due to the lane blockage being before/after on-ramp/off-ramp (compared to the density on a straight stretch of roadway) is considered to be the

same for either blockage. The effective capacity of the bottleneck is calculated taking the lane blockage into consideration.

- **Two lane blockage** : Both the lanes in the tunnel are blocked in this case. The effective capacity is reduced to zero and the vehicles get stranded before the incident location.

In the following sections, each of the above cases is considered in more detail and a control strategy suggested. Due to the absence of a CO model that gives the exact CO exposure of each vehicle at any instant of time, the following assumptions are made:

- A density lesser than or equal to $D_{Critical}$ on a link is of no concern with regards to CO levels on that link.
- A density greater than $D_{Critical}$ on a link implies that the vehicles on the link have crossed the critical CO level exposures and precautions have to be taken to ensure that no more vehicles get affected in this manner.

The sections are defined based on the location of the VSLS. The zone of influence of each VSLS forms a section. If the link/segment on which the density is found high falls under more than one section, then the same control actions are taken on both these sections.

No lane blockage

The cause of alarm is the density on one of the links in the tunnel, which is found to be high (greater than $D_{Critical}$). Consider the scenario in Figure C-13.
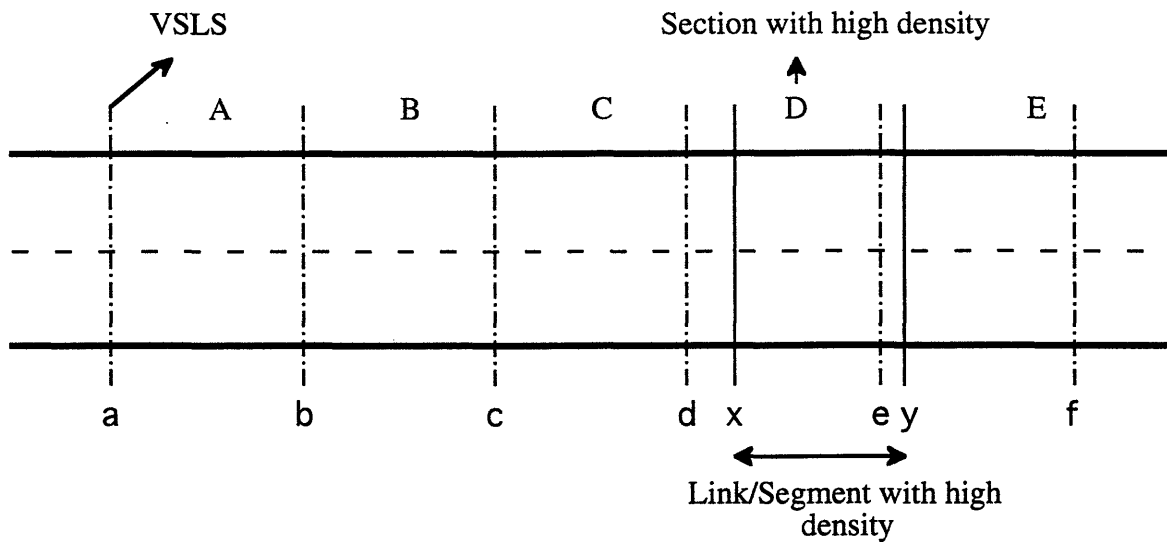
Figure C-13 A high density scenario

Let x and y mark the beginning and end of the link/segment detected with high density. As we have said before, the sections are marked based on the zones of influence of the VSLS. Therefore, for implementing the speed limit signs, the sections on which the densities are high need to be known. Based on the position of x and y relative to their respective nearest VSLS, we decide whether the sections to which x and y belong, are affected by high density.

If $(dx > X_{hd})$ and $(ey < X_{hd})$, it means that the link/segment is too small to be controlled for density.

If $(dx > X_{hd})$ and $(ey > X_{hd})$, then only section E is the section with high density.

If $(dx < X_{hd})$ and $(ey < X_{hd})$, then only section D is the section with high density.

If $(dx < X_{hd})$ and $(ey > X_{hd})$, then it means that both the sections D and E are affected by high density.

If the segment on which the high density is detected covers any VSLS section between x and y completely, then these VSLS sections are declared to be of high density and the speed limits on these sections are based on observed traffic conditions.

Assume in the scenario considered that $dx < X_{hd}$ and $ey < X_{hd}$, so that only section D is the section with high density. One way to reduce the link density is to decrease the inflow to the link and increase the outflow from the link. The latter can be done by increasing the speed limits on the sections downstream of the affected link. The speeds are initially observed on the downstream section (Say, S, when rounded off to the higher multiple of 5) and the speed limit is set to be $(S + S_{addds})$. The inflow to the link can be

153

reduced by postponing the further arrival of vehicles at the beginning of the link. An increasing speed gradient is created along the flow direction, using speed limit signs, which delays the upstream vehicles and clears off the high density downstream. This is done until the CO levels return to normal (density should be at most $D_{Critical}$), when the speed limits can again be changed to the normal limits, based on the observed traffic conditions.

Figure C-14 shows a simple way of creating a speed gradient along the tunnel section. Upstream of B and downstream of E, the speed limits are based on the speeds obtained from the observed traffic conditions.
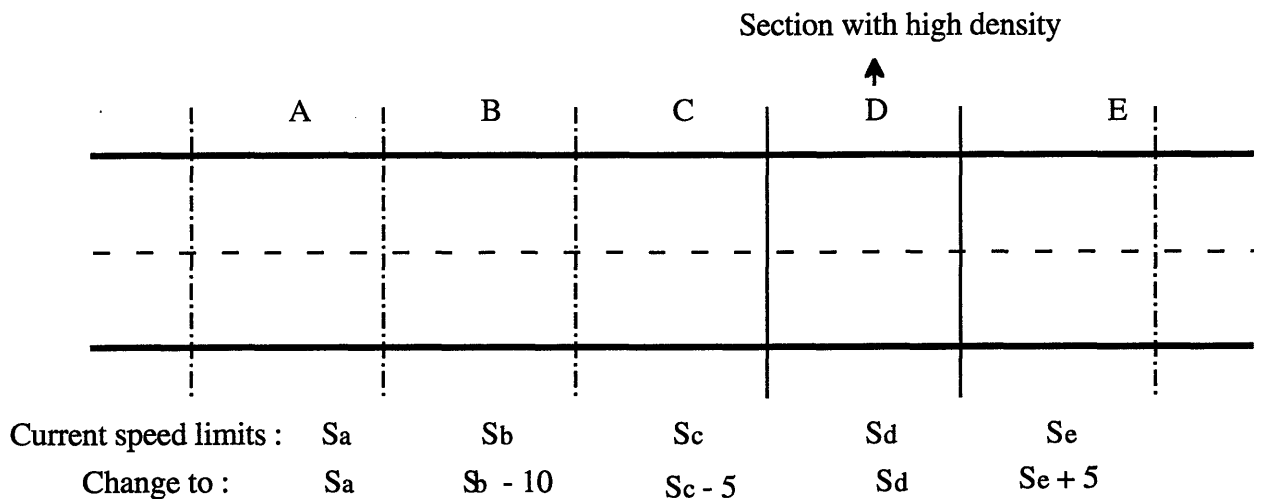
Section with high density

Current speed limits :	Sa	Sb	Sc	Sd	Se
Change to :	Sa	Sb - 10	Sc - 5	Sd	Se + 5

Figure C-14   Relative changes to speed limits on different sections

One consequence of the above strategy is an increased density on the upstream links due to reduced speed limits. If these links are also part of the tunnel, then, the result of the control strategy is only the shifting of the problem location. To prevent this from happening, the portal signals at the tunnel entrance are used. Therefore, if the density on a new link upstream of the original high density link crosses the critical limits, then the following strategy is used.

The density desired in the tunnel to get the CO levels to within safety limits is given as $D_{Desired} = \xi * D_{Critical}$, where ξ is the 'desired density ratio'. The flow rate needed to maintain this density is calculated using the flow-density relationship for that link. Assume that the flow rate desired is $F_{Desired}$ (vph). The metering at portal cannot be

154

done independent of the ramps near the affected link. The presence of any on-ramps to the tunnel section and their contribution to the demand at the bottleneck section should be taken into consideration while calculating the metering rate. Since the ramps cannot be completely closed-off, a minimum metering rate of 240 vph (4 vpm) is used at the on-ramps. It is assumed that only on-ramps within 1000 ft upstream of the affected section will affect the flow at the bottleneck. Therefore, if there is an on-ramp within 1000 ft upstream of section D, the metering rate at the portal is ($F_{Desired}$ - 240) vph which gives ($F_{Desired}$ - 240)/2 vph per lane. The number of cycles per minute and the lengths of the green, yellow and red times in each cycle so as to maintain the required flow rate needs to be decided. The following rules are used in this decision process:

- The metering is done with at least 2 cycles and atmost 4 cycles every minute.
- To obtain the green time in a cycle, assume that each vehicle takes 3 seconds to proceed past the portal signal. After allowing for all the required number of vehicles in the cycle to pass, the left over time in the cycle is used for red. If the red time falls below 3 seconds, then 3 seconds is allotted for red and the signal is kept green for the remaining time of the cycle. The metering at the portal is stopped when the density on all the links in the tunnel drops to less than or equal to $D_{Desired}$.

With regards to the speed limits, creating a speed gradient does not make sense because the access to the tunnel is controlled anyway. Let the desired speed obtained from substituting the desired density in the speed-density relationship for the link be denoted by $S_{Desired}$. This desired speed is imposed on all the sections upstream of the bottleneck section until the portal is encountered. Upstream of the portal, a speed gradient as suggested in the first strategy is used to ensure that there is no queue formation even outside the queue. This also ensures that there is no sudden drop of speeds from outside the tunnel to inside. These speed limits are used only during the time the portal metering is done. When the metering is stopped, the speed limits are again based on the observed traffic conditions.

One lane blockage

The control strategy is similar to the one for 'no lane blockage' except that if portal signals are used, the metering rates are decided based on the effective capacity of the bottleneck. The bottleneck section capacity is computed taking the lane closure into consideration.

Let the bottleneck capacity be C. If $F_{Desired}$ is greater than C, then we change $F_{Desired}$ to C, since we do not want any queue buildup inside the tunnel, which again results in increased densities. If $F_{Desired}$ is less than C, then we use $F_{Desired}$ as the flow rate since we want to adhere by the CO limits.

Two lane blockage

As mentioned before, when both the lanes are blocked, the effective capacity becomes zero. All the vehicles upstream of the blockage and already in the tunnel cannot leave the tunnel and hence will come to a halt before the blockage. These vehicles are advised to shut their motors after stopping. To prevent any more vehicles from entering the tunnel upstream of the blockage, all the access points (entrance and ramps) to the tunnel upstream of the blockage are closed. The tunnel entrance is closed using portal signals. These access points are reopened only when the incident is cleared and normal conditions are restored.

The pseudo code of the control strategy for 'reducing link density' is given below:

No lane blockage

Identify the link with high density
    Identify the section with high density (D)
        Increase the outflow from the section
            Increase the speed limit on section immediately downstream of D.
        Decrease the inflow to the section
            Create a speed gradient by reducing the speed limits on upstream
                sections.
    If this increases density on upstream links,
        Use portal signals to control flow into the tunnel.
            Let $D_{Desired} = \xi * D_{Critical}$.
                Obtain corresponding flow from the flow-density relationship ($F_{Desired}$)
                Use a minimum metering rate of 240 vph on on-ramps.
                If there is an on-ramp within 1000 ft upstream of the beginning of the high
                    density section, metering rate at portal = ($F_{Desired}$ - 240) vph.
                Stop metering when density becomes $\leq D_{Desired}$

Obtain desired speed, $S_{Desired}$ corresponding to $D_{Desired}$.

Use $S_{Desired}$ as speed limit on all sections in tunnel upstream of high density link.

When metering is stopped,

Speed limits are based on observed traffic conditions.

One lane blockage

Everything same as for 'no lane blockage' except that

Metering rates at portal are decided based on the bottleneck capacity (C).

If ($F_{Desired} \geq C$),

$F_{Desired} = C$.

Else

Use $F_{Desired}$ as desired flow rate.

Two lane blockage

Close all access points to the tunnel upstream of the blockage location

Appendix D Scenarios for Case Study

In this appendix, a description of the demand patterns, event patterns and behavioral parameter sets that are used in generating scenarios is given.

D.1 Traffic Demand

The network has two origin nodes: 1, 2 and two destination nodes: 5, 6. Therefore, four origin-destination (O-D) pairs are possible: (1,5), (1,6), (2,5), (2,6). The simulations are of 30 minutes each and the O-D matrix is specified for every 10 minute intervals. PM peak hourly volume of 3570 vph is used as the basis. Changes are made to this volume for every 10 minute interval in each demand pattern. The vehicle mix used is specified in Table D-1.

Table D-1 Vehicle mix used for case study

Vehicle Type	Percentage
SOV (Cars)	86%
HOV (Buses)	10%
Trucks	4%

The split among the 4 OD pairs, given the link volume of (3,4) is given in Table D-2.

Table D-2 Vehicle split among O-D pairs for case study

O-D Pair	Percentage of peak hourly demand
(1,5)	43.5%
(1,6)	25.5%
(2,5)	19.5%
(2,6)	11.5%

Two demand patterns are specified for implementation. In the *first pattern*, the demand is kept almost at the observed peak volumes, on the average. The first demand pattern is shown in Figure D-1 and the corresponding O-D matrix is shown in Table D-3.
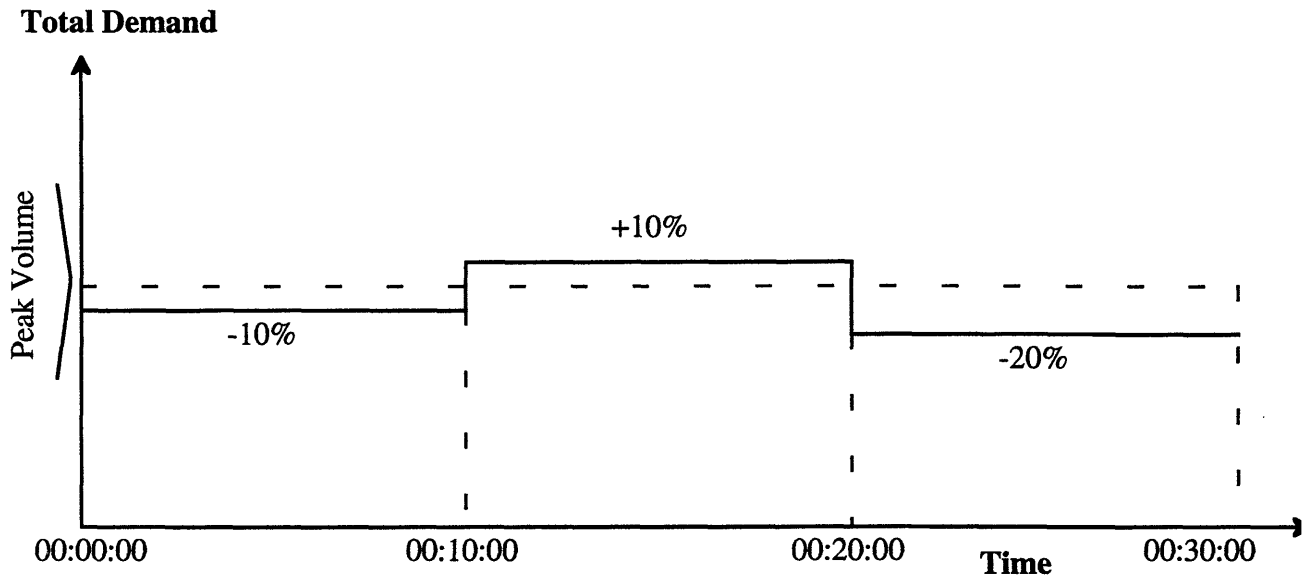
Total Demand

Figure D-1 Pictorial representation of the first traffic demand pattern

Table D-3 O-D Matrix for the first traffic demand pattern

00 : 00 : 00 to 00 : 10: 00			
OD Pair	Vehicles per hour of each type		
	Cars	Buses	Trucks
(1,5)	1202	140	56
(1,6)	704	82	33
(2,5)	538	63	25
(2,6)	318	37	15
00 : 10 : 00 to 00 : 20: 00			
OD Pair	Vehicles per hour of each type		
	Cars	Buses	Trucks
(1,5)	1469	171	68
(1,6)	861	100	40
(2,5)	659	76	31

159

(2,6)	389	45	18
00 : 20 : 00 to 00 : 30: 00			
OD Pair	Vehicles per hour of each type		
	Cars	Buses	Trucks
(1,5)	1068	124	50
(1,6)	626	73	29
(2,5)	479	56	22
(2,6)	283	33	13

In the *second pattern*, the demand is taken to be much higher than the observed peak volumes. The second demand pattern is shown in Figure D-2 and the corresponding O-D matrix is shown in Table D-4.
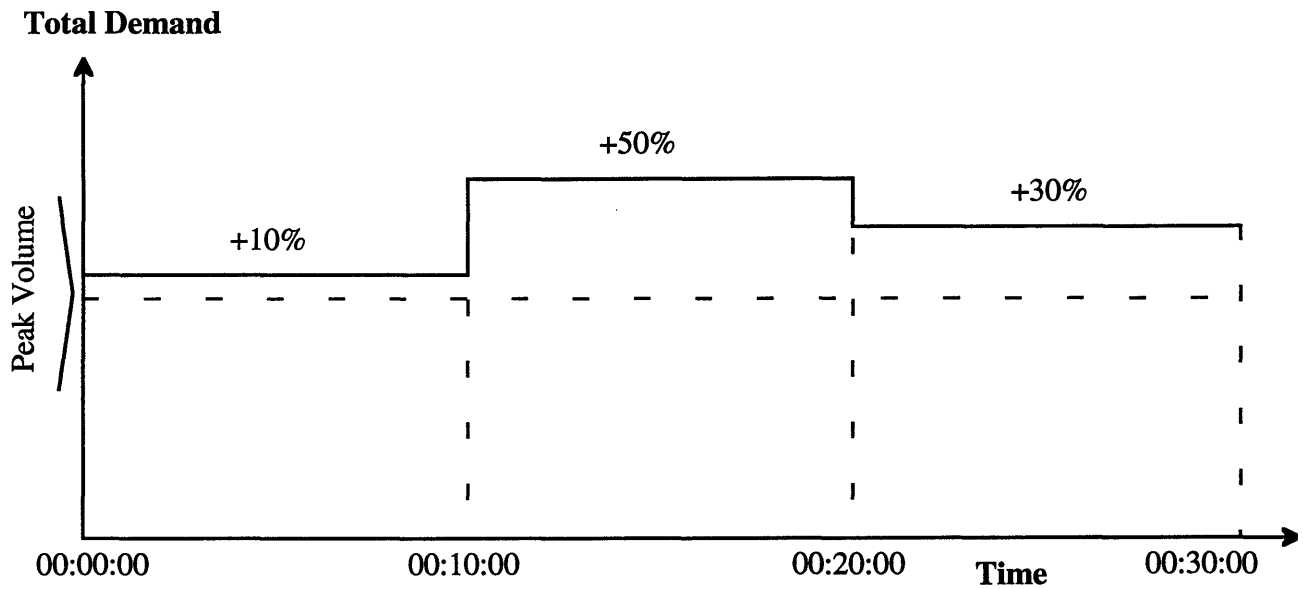

Figure D-2 Pictorial Representation of the second traffic demand pattern

Table D-4 O-D Matrix for the second traffic demand pattern

00 : 00 : 00 to 00 : 10: 00			
OD Pair	Vehicles per hour of each type		
	Cars	Buses	Trucks
(1,5)	1469	171	68

(1,6)	861	100	40
(2,5)	659	76	31
(2,6)	389	45	18

00 : 10 : 00 to 00 : 20: 00			
OD Pair	Vehicles per hour of each type		
	Cars	Buses	Trucks
(1,5)	2003	233	93
(1,6)	1175	136	55
(2,5)	898	104	42
(2,6)	530	61	25

00 : 20 : 00 to 00 : 30: 00			
OD Pair	Vehicles per hour of each type		
	Cars	Buses	Trucks
(1,5)	1736	202	81
(1,6)	1017	119	47
(2,5)	778	91	36
(2,6)	459	54	21

D.2 Event Patterns

For the event pattern, the events are pre specified. They are not generated in real-time as described in Chapter 3. Two events are specified for use in the simulation. If the scenarios were generated iteratively, then one event pattern would have been generated initially. However, this is not the case for the case study, hence, two events are specified.

The first pattern has two incidents, both occurring on the mainline one after another with a gap of 4.5 minutes. Figure D-3 illustrates the relative location of the incidents and gives the details of the incidents.
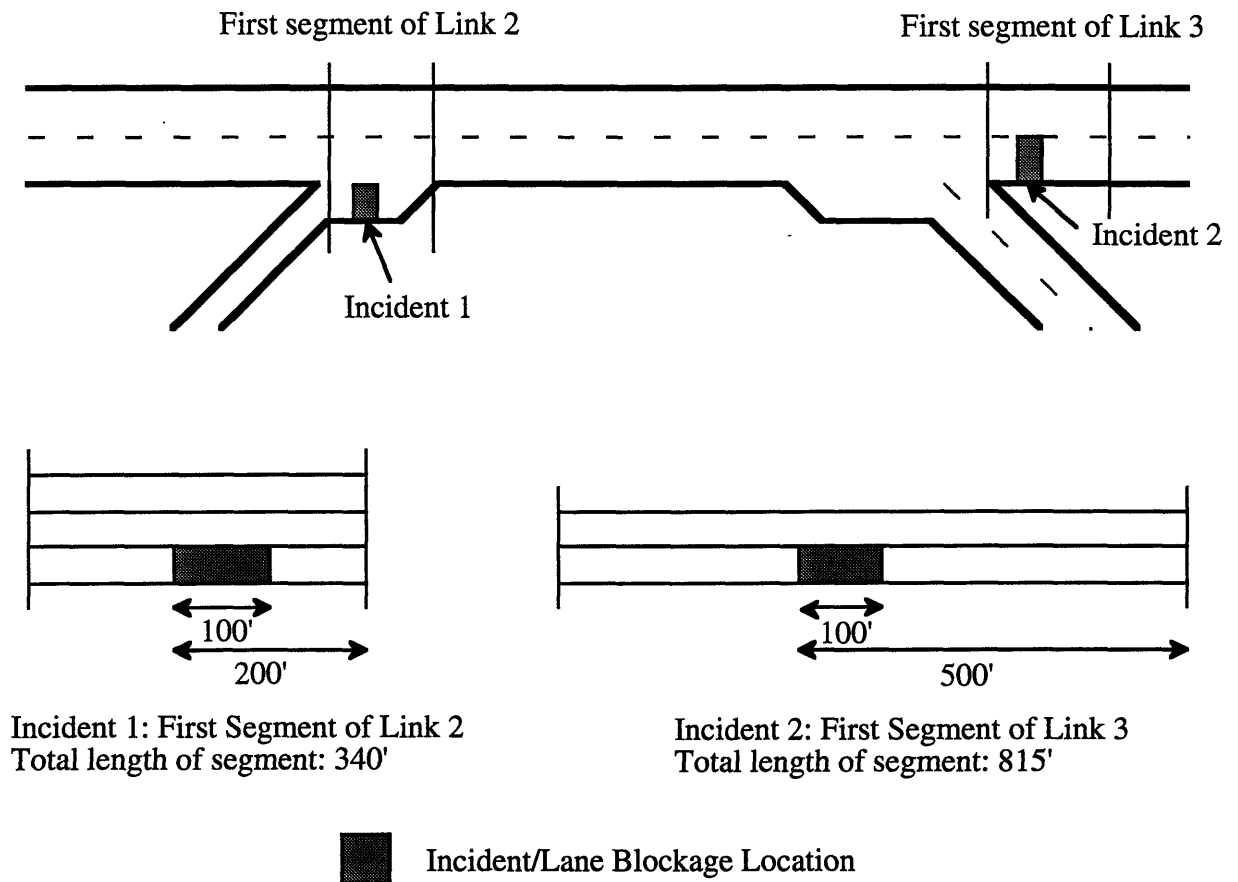
Figure D-3 First event pattern for the case study
Figure not to scale

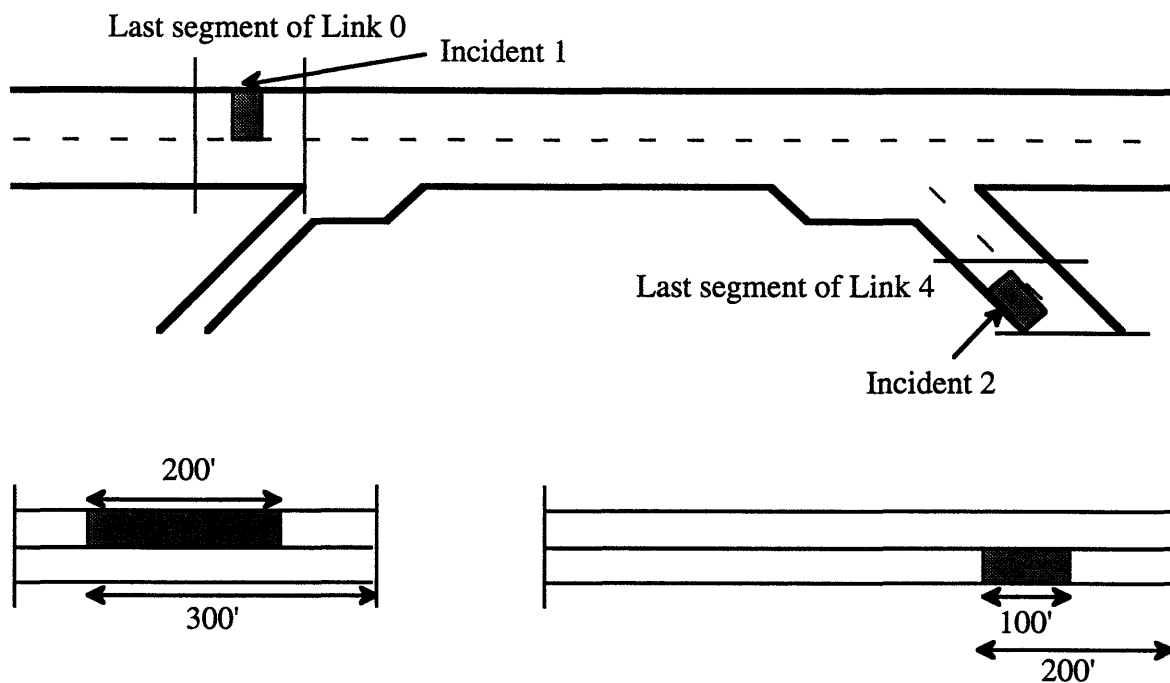The incident 1 has the following characteristics:

- Single right lane blockage
- Occurs after the on-ramp completely merges with the mainline
- Occurs on link 2, first segment
- Length of blockage: 100 ft
- The beginning of the blockage is located 200 ft from the end of the segment
- Occurs at 02:00
- Event duration: 8:30 (Therefore, it lasts till 10:30)

The incident 2 has the following characteristics:

- Single right lane blockage
- Occurs after the off-ramp
- Occurs on link 3, first segment

- Length of blockage: 100 ft
- The beginning of the blockage is located 500 ft from the end of the segment
- Occurs at 17:00
- Event duration: 5:00 (Therefore, it lasts till 23:00)

The second pattern also has two incidents, one occurring on the mainline and the other on the off-ramp. The second incident takes place (begins and ends) while the first one is still active. Figure D-4 illustrates the relative location of the incidents and gives the details of the incidents.

Incident 1: Last Segment of Link 0
Total length of segment: 375'

Incident 2: Last Segment of Link 4
Total length of segment: 700'

Incident/Lane Blockage Location

Figure D-4 Second event pattern for the case study
Figure not to scale

1. The incident 1 has the following characteristics:
   - Single left lane blockage
   - Occurs before the on-ramp

163

- Occurs on link 0, last segment
- Length of blockage: 200 ft
- The beginning of the blockage is located 300 ft from the end of the segment
- Occurs at 06:00
- Event duration: 12:00 (Therefore, it lasts till 18:00)

2. The incident 2 has the following characteristics:
- Single lane blockage
- Occurs on the off-ramp
- Occurs on link 4, last segment
- Length of blockage: 100 ft
- The beginning of the blockage is located 200 ft from the end of the segment
- Occurs at 08:00
- Event duration: 7:00 (Therefore, it lasts till 15:00)

D.3 Behavioral Parameters

These parameters capture the probabilistic behavior of the driver. It was mentioned in section 3.4 that most of the driver behavior variables are specified as distributions. However, for simplicity, a deterministic version of this has been used for the case study. Only the desired speed ratio, merging headway buffer, and startup delay are drawn from distributions. The parameters that capture the behavior relate to:
- Desired Speed Ratio
- Car Following Parameters : H^{upper}, H^{lower}, α, β, γ, Δt
- Lane Changing Parameters : ρ, δ, ζ, h^{lead}, h^{lag}
- Probability of courtesy yielding : $p^{DY/MY}$
- Merging Headway Buffer
- Startup Delay

Parameter set 1

All the parameters assume nominal values. This gives nominal behavior of the drivers. The nominal values of the parameters are given below:

- Desired Speed Ratio Distribution:

Driver Population	Accumulated Population	Desired Speed Ratio
0.05	0.05	0.80
0.05	0.10	0.90
0.25	0.35	1.00
0.35	0.70	1.10
0.20	0.90	1.20
0.10	1.00	1.30

- Car Following Parameters :

$\{H^{upper} : 4.0, H^{lower} : 0.5, \alpha : 2.0, \beta : 1.0, \gamma : 1.0, \Delta t : 1.5\}$

- Lane Changing Parameters :

$\{\rho : 3.85*10^{-3}, \delta : 2.0, \zeta : 0.9,$

$(h^{lead}, h^{lag}) : (1.0, 2.0)$ for discretionary lane changes and

$(0.5, 2.0)$ for mandatory lane changes$\}$

- Probability of courtesy yielding : $p^{DY/MY} : 0.5$
- Merging Headway Buffer : Uniform distribution between 3.0 and 6.0 seconds
- Startup Delay : Uniform distribution between 0.0 and 3.0 seconds.

Parameter set 2

The drivers are made aggressive by applying the following changes:

Decrease the value of H^{upper} : 2.5

Decrease the value of H^{lower} : 0.3

Increase the value of ρ : $7.70*10^{-3}$

Decrease the value of δ : 1.0

Decrease the values of h^{lead} and h^{lag}

      Discretionary lane changing : (0.5, 1.0)

      Mandatory lane changing : (0.5, 1.0)

Decrease the value of $p^{DY/MY}$: 0.1

Decrease the value of Merging Headway Buffer : Uniform distribution between 1.5 and 3.0 seconds

Decrease the value of Startup Delay : Uniform distribution between 0.0 and 0.5 seconds.

All the remaining parameters are kept at their values in parameter set 1.

Glossary of Abbreviations

APTS	Advanced Public Transportation Systems
ARTS	Advanced Rural Transportation Systems
ATIS	Advanced Traveler Information Systems
ATMS	Advanced Traffic Management Systems
AVCS	Advanced Vehicle Control Systems
AVI	Auto Vehicle Identification
CA/T	Central Artery/Tunnel
CCTV	Closed Circuit TeleVision
CO	Carbon Monoxide
CVO	Commercial Vehicle Operations
DRL	Downstream Red Length
DTMS	Dynamic Traffic Management Systems
HAR	Highway Advisory Radio
HOV	High Occupancy Vehicles
IID	In-vehicle Information Devices
IPCS	Integrated Project Control System
LOV	Low Occupancy Vehicles
LUS	Lane Use Sign
MIT	Massachusetts Institute of Technology
MITSIM	MIcroscopic Traffic SIMulator
TMC	Traffic Management Center
UYL	Upstream Yellow Length
URL	Upstream Red Length
VMS	Variable Message Sign
VRC	Vehicle to Roadside Communication
VSLS	Variable Speed Limit Sign

Bibliography

[Ben-Akiva 92] Moshe E. Ben-Akiva, David Bernstein, Anthony F. Hotz, Harilaos N. Koutsopoulos, and Joseph M. Sussman. (1992). "The case for smart highways". Technology Review, July, pp. 39-47.

[Ben-Akiva 94] Moshe E. Ben-Akiva, H. N. Koutsopoulos, R. G. Mishalani, Q. Yang. (1994). "Integrated simulation framework for evaluating dynamic traffic management systems". Paper accepted for publication in the proceedings of the First World Congress on Applications of Transport Telematic and Intelligent Vehicle Highway Systems, Paris, France, 1994.

[Cassidy 92] Michael J. Cassidy and Kumares C. Sinha. (1992). "An application of advanced technologies for freeway traffic management - An Indiana case study". Proceedings of the Second International Conference on Applications of Advanced Technologies in Transportation Engineering.

[CATHT 90] Central Artery (I-93)/Third Harbor Tunnel (I-90) project, Boston, Massachusetts : Final supplemental environmental impact statement/report, EOEA #4325 / submitted by the Massachusetts Department of Public Works; prepared by Bechtel/Parsons Brinckerhoff. (1990)

[EPA/FHWA 93] EPA/FHWA CO Guidance. (1993). Personal communication with R. Brindley, Massachusetts Highway Department, Central Artery/Tunnel.

[Gartner 92] Nathan H. Gartner and Dennis L. Hou. (1992). "Comparative evaluation of alternative traffic control strategies". TRR 1360.

[Giuliano 85] Genevieve Giuliano. (1985) "A multicriteria method for transportation investment planning". Transportation Research A, Vol. 19A, No. 1, pp. 29-41.

[Head 1992] Larry K. Head, Pitu B.Mirchindani and Dennis Sheppard. (1992). "Hierarchical framework for real-time traffic control". TRR 1360

[Hobieka 92]   A. G. Hobieka and Y. Zhang. (1992). "A real-time traffic diversion model: A conceptual approach". Proceedings of the Second International Conference on Applications of Advanced Technologies in Transportation Engineering.

[Hotz 94]   A. F. Hotz, H. N. Koutsopoulos, M. Ben-Akiva, D. Bernstein, A. Chachich, R. Mishalani, N. V. Jonnalagadda, Q. Yang, B. Brindley, D. Krechmer, C.T. Marcus. (1994). "Evaluating traffic control systems using microsimulation: The Central Artery/Tunnel Integrated Project Control System". Paper presented at IVHS America Fourth Annual Meeting, Atlanta, Georgia.

[HS 90]   Highway Statistics (1990)

[IPCSDR 93]   "Integrated Project Control System Design Refinements". (1993). Report submitted by the MIT IVHS Research Program to Bechtel/Parsons Brinckerhoff.

[IVHSMS 93]   "Technology for the 21st Century", Mini-symposium on Intelligent Vehicle-Highway Systems, March 13, 1993, GTE Laboratories Incorporated, Waltham, Massachusetts.

[Kroes 83]   Kroes J. L. de, Donk P. and Klien S. J. de. (1983). Evaluatie van de externe effecten van het VerkeersSignalerings-Systeem voor Autosnelwegen. Report.

[Kwon 92]   Eil Kwon. (1992). "A new approach for real-time prediction of traffic demand-diversion in freeway corridors". Proceedings of the Second International Conference on Applications of Advanced Technologies in Transportation Engineering.

[Meyer 84]   Michael D. Meyer and Eric J. Miller. (1984) "Urban Transportation Planning : A Decision Oriented Approach". McGraw-Hill Publishers.

[Montgom]   Douglas C. Montgomery. (1991). "Design and Analysis of Experiments", 3rd edition. Wiley Publishers.

[Mukundan 92] Anil Mukundan. (1992) "A real-time system for IVHS traffic modeling". Master's thesis, Massachusetts Institute of Technology, Cambridge, MA.

[Pardo 93]   O. Ray Pardo. (1993). "Advanced highway control on the Central Artery". Proceedings of the ASCE Third International Conference on Applications of Advanced Technologies in Transportation Engineering.

[Parkany 93] A. Emily Parkany. (1993) "Using Vehicle to Roadside Communication data for incident detection". Master's Thesis, Massachusetts Institute of Technology, Cambridge, MA.

[Reiss 91]   Robert A. Reiss, Nathan H. Gartner and Stephen L. Cohen. (1991). "Dynamic control and traffic performance in a freeway corridor: A simulation study". Transportation Research A, Vol. 25A, No. 5, pp 267-276.

[Rietveld 80] Rietveld P. (1980). "Multiple objective decision methods and regional planning". Studies in Regional Science and Urban Economics, Vol. 7. North-Holland, Amsterdam. North-Holland Publishing Company.

[Rilett 92]   L. Rilett and M. Van Aerde. (1992). "Routing based on anticipated travel times". Proceedings of the Second International Conference on Applications of Advanced Technologies in Transportation Engineering.

[Roy 81]   Roy B. and Vincke P. (1981). "Multicriteria analysis: Survey and new directions". European Journal of Operational Research 8, pp 207-218.

[Shepherd 93] Simon P. Shepherd. (1993). "Metering strategies applied to signalized networks". Proceedings of the ASCE Third International Conference on Applications of Advanced Technologies in Transportation Engineering.

[Smith 92]   Steven A. Smith and Cesar Perez. (1992). "Evaluation of INFORM: Lessons learned and application to other systems". TRR 1360.

[Smulders 90] Stef Smulders. (1990). "Control of freeway traffic flow by variable speed signs". Transportation Research B, Vol. 24B, No. 2, pp. 111-132.

[SPIVHS 92] "Strategic plan for Intelligent Vehicle-Highway Systems in the United States" (1992). Report No. IVHS-AMER-92-3. Prepared by IVHS AMERICA.

[Stephan 89] Yorgos J. Stephanedes, Eil Kwon, and Michalopoulos, P.G. (1989). "Demand diversion for vehicle guidance, simulation and control in freeway corridors". TRR 1220.

[Stephan 92-1] Yorgos J. Stephanedes and Kai-Kuo Chang. (1992). "Optimal ramp-metering control for freeway corridors". Proceedings of the Second International Conference on Applications of Advanced Technologies in Transportation Engineering.

[Stephan 92-2] Yorgos J. Stephanedes, Eil Kwon, and Kai-Kuo Chang. (1992). "Control emulation method for evaluating and improving traffic-responsive ramp metering strategies". TRR 1360.

[Sussman 93] Joseph M. Sussman. (1993). "Intelligent Vehicle Highway Systems". Construction Business Review, May/June.

[TCSH 76] "Traffic control systems handbook" (1976). US DOT. FHWA.

[TCSH 85] "Traffic control systems handbook- Revised" (1985). US DOT. FHWA.

[Yang 93] Qi Yang. (1993). "A microscopic traffic simulation model for IVHS applications". Master's Thesis, Massachusetts Institute of Technology, Cambridge, MA.