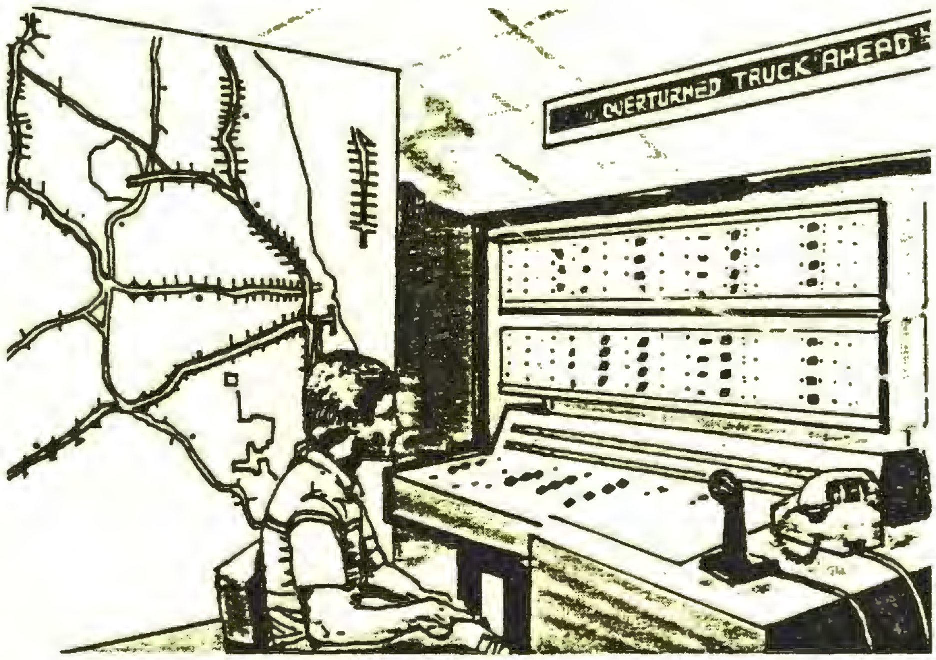
25 dup

Prepared For
Federal Highway Administration

U.S. Department of Transportation

A Freeway Management Handbook



Volume 2: Planning & Design

Volume 1: Overview
Volume 2: Planning & Design
Volume 3: Operations & Maintenance
Volume 4: Annotated Bibliography

May 1983

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Freeway Management Handbook, Volume 2, Planning and Design		5. Report Date May 1983 6. Performing Organization Code
7. Author(s) R. Sumner, S. Shapiro, D. Capelle, D. Hill P. Tarnoff, J. Petrykanyn, J. Watt		8. Performing Organization Report No.
9. Performing Organization Name and Address PRC Voorhees 1500 Planning Research Drive McLean, Virginia 22102	discarded	10. Work Unit No. (TRAIS) 11. Contract or Grant No. DOT-FH-11-9706
12. Sponsoring Agency Name and Address Federal Highway Administration U.S. Department of Transportation Washington, D.C. 20590		13. Type of Report and Period Covered Final Report April 1981 — May 1983 14. Sponsoring Agency Code

15. Supplementary Notes

FHWA Contract Manager: E. Munley

16. Abstract

Freeway Managment can be defined as the control, guidance and warning of traffic in order to improve the flow of people and goods on these limited access facilities. This handbook has been designed to provide potential users with a set of guidelines for the planning, design, operation, and maintenance of the various components of Freeway Management systems. The topics that are addressed in the handbook include:

- Problem Identification and Analysis
- Ramp Metering
- Priority Treatments for High Occupancy Vehicles
- Incident Detection
- Organization of Incident Response
- Driver Information Systems
- Evaluation of alternative Improvements
- Use of Simulation Models
- Implementation Issues
- Surveillance Operations
- Management of Traffic Operations
- Selection and Maintenance of System Components

This volume is aimed principally at engineers. It is designed to enable engineers to analyze the performance of existing freeway systems and gives instructions on how various proposed strategies can be evaluated. By quantifying the benefits of a range of potential solutions, the engineer using volume 2 will be directed toward the most viable technique.

Volume 1 — Overview
Volume 2 — Planning and Design
Volume 3 — Operations and Maintenance
Volume 4 — Annotated Bibliography

17. Key Words		18. Distribution Statement
Freeway Management Freeway Operations Ramp Metering Incident Detection Incident Management	Priority Lanes Motorist Information Systems Traffic Surveillance Systems	

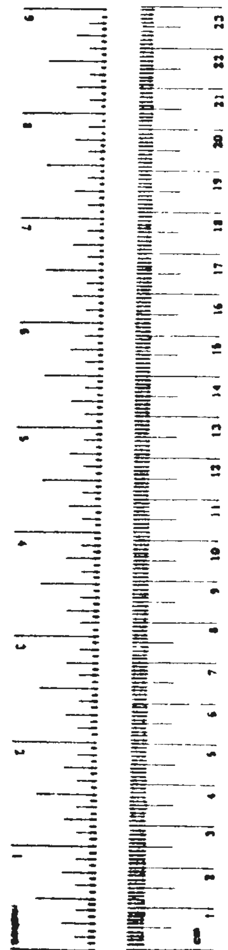
19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	220	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

METRIC CONVERSION FACTORS

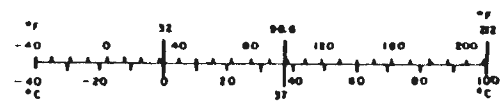
Approximate Conversions to Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
in	inches	2.5	centimeters	cm
ft	feet	30	centimeters	cm
yd	yards	0.9	meters	m
mi	miles	1.6	kilometers	km
		AREA		
in^2	square inches	6.5	square centimeters	cm^2
ft^2	square feet	0.09	square meters	m^2
yd^2	square yards	0.8	square meters	m^2
mi^2	square miles	2.6	square kilometers	km^2
	acres	0.4	hectares	ha
		MASS (weight)		
oz	ounces	28	grams	g
lb	pounds	0.45	kilograms	kg
	short tons (2000 lb)	0.9	tonnes	t
		VOLUME		
tsp	teaspoons	5	milliliters	ml
Tbsp	tablespoons	15	milliliters	ml
fl oz	fluid ounces	30	milliliters	ml
c	cups	0.24	liters	l
pt	pints	0.47	liters	l
qt	quarts	0.95	liters	l
gal	gallons	3.8	liters	l
ft^3	cubic feet	0.03	cubic meters	m^3
yd^3	cubic yards	0.76	cubic meters	m^3
		TEMPERATURE (exact)		
°F	Fahrenheit temperature	5/9 (after subtracting 32)	Celsius temperature	°C



Approximate Conversions from Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
mm	millimeters	0.04	inches	in
cm	centimeters	0.4	inches	in
m	meters	3.3	feet	ft
m	meters	1.1	yards	yd
km	kilometers	0.6	miles	mi
		AREA		
cm^2	square centimeters	0.16	square inches	in^2
m^2	square meters	1.2	square yards	yd^2
km^2	square kilometers	0.4	square miles	mi^2
ha	hectares (10,000 m^2)	2.5	acres	
		MASS (weight)		
g	grams	0.035	ounces	oz
kg	kilograms	2.2	pounds	lb
t	tonnes (1000 kg)	1.1	short tons	
		VOLUME		
ml	milliliters	0.03	fluid ounces	fl oz
l	liters	2.1	pints	pt
l	liters	1.06	quarts	qt
l	liters	0.26	gallons	gal
m^3	cubic meters	35	cubic feet	ft^3
m^3	cubic meters	1.3	cubic yards	yd^3
		TEMPERATURE (exact)		
°C	Celsius temperature	9/5 (then add 32)	Fahrenheit temperature	°F



A FREEWAY MANAGEMENT HANDBOOK

VOLUME 2: PLANNING AND DESIGN

Prepared for

FEDERAL HIGHWAY ADMINISTRATION

by

PRC VOORHEES
1500 Planning Research Drive
McLean, Virginia 22102

May 1983

TABLE OF CONTENTS

Table of Contents, Continued

LIST OF FIGURES

List of Figures, Continued

List of Figures, Continued

LIST OF TABLES

List of Tables, Continued

1. INTRODUCTION

HOW TO USE THIS HANDBOOK

This is volume 2 of a four-volume handbook designed to provide information on the efficient management of freeways. This volume has been written for engineers and will provide a better understanding of how to analyze and design new freeway management systems, and how to modify existing systems.

Volume 1 is designed principally for policymakers and administrators, and provides a broad perspective of freeway management. Volume 3 deals with implementation and is aimed at engineers, technicians, and emergency services personnel. Volume 3 should be used when information on construction techniques or installation procedures is required. Volume 4 provides an annotated bibliography with a cross-referencing index that enables readers needing more information to find references to other freeway management publications.

The four sections of this volume are: "Identifying and Analyzing Problems," "Designing Solutions," "Implementing Solutions," and "Evaluating Performance." Each section is preceded by a "section guide" that gives a brief description of the type of information contained in each chapter.

Identifying and Analyzing Problems

The various techniques for identifying and analyzing problems are discussed in chapters 2 through 5; chapter 3 is the core of this section. The procedures for analyzing the various alternatives that are contained in this chapter can also be used for the evaluation of existing conditions. The reader going through this material for the first time may wish to review chapter 3 first (bearing in mind that the analysis can be simplified using actual values measured in the field) and then refer to the other chapters in this section for additional detail.

Designing Solutions

Chapter 13 is the core of the "Designing Solutions" section. It contains a procedure for evaluating solutions that can be applied to both recurring and non-recurring congestion problems, and identifies the pros and cons of different cost-effectiveness evaluation techniques. The first-time reader may wish to concentrate on chapter 13 and then skim the other chapters, reading them in more detail as the various stages of the analysis progress.

Implementing Solutions

The third section, "Implementing Solutions" discusses the implementation issue of financial planning, staged construction, and interagency cooperation in two brief chapters.

Evaluating Performance

The last section is an epilogue that describes the importance of a performance evaluation and how this evaluation can be planned and conducted.

Many aspects of freeway management overlap in their subject areas. Readers interested in a specific chapter may wish to read other chapters that are closely related to their principal area of interest. Figure 1.1 is a chart that can be used as a guide to locate these related chapters. The reader should look at the column marked Principal Chapter of Interest, Supplementary Reading

		Principal Chapter of Interest	Vol. 2 Planning & Design: 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Vol. 3 Operations and Maintenance: 2	3	4	5	6	7	8	9	10	11	12	13	14	15
Volume 2 Planning and Design	Identifying and Analyzing Problems	2 Data Collection	■				●						●					●	●						●			●			
		3 Delay Computation Methods		■	●			●						●				●	●												
		4 Environmental Impacts	●		■					●				●								●									
		5 Simulation Models in Diagnosis and Prediction				■						●			●																
	Designing Solutions	6 Detection of Recurring Congestion	●			●	■						●	●				●	●	●					●			●			
		7 Ramp Control	●	●				■											●	●					●			●	●	●	●
		8 Mainline and Corridor Control	●	●					■										●	●					●			●	●	●	●
		9 Bus/Carpool	●	●						■									●	●					●			●	●	●	●
		10 Detection of Non-Recurring Congestion									■	●	●				●								●		●	●			
		11 Incident Detection Algorithms										■	●																	●	
		12 Planning for Incident Response		●		●						●	■		●		●				●	●			●	●	●	●			
		13 Cost-Effectiveness Evaluation Techniques				●								■		●															
		14 Simulation Models in Design and Evaluation				●									■																
	Implementing Solutions	15 Staged Construction and Financial Planning														■															●
		16 Interagency Cooperation											●				■				●	●		●							
Volume 3 Operations and Maintenance	Surveillance Operations	2 Data Gathering for Continuous Evaluation	●	●			●						●					■	●						●			●	●	●	●
		3 Detection of Congestion	●				●						●						■	●					●			●	●		
	Managing Traffic Operations	4 Managing Recurring Congestion					●	●	●	●										■							●				
		5 Coordinating Incident Response															●				■	●	●	●			●				
		6 On-Site Incident Management															●					■	●	●							
		7 Dispatchers' Manual	●														●						■	●			●			●	
		8 Hazardous Materials Manual															●				●	●	●	■							
	Selecting and Maintaining System Components	9 System Monitoring and Maintenance	●	●			●											●							■						●
		10 Motorist Information System											●													■			●	●	●
		11 Closed-Circuit TV					●						●								●	●					■		●	●	●
		12 Traffic Detectors	●				●	●	●				●					●	●						●			■	●		●
		13 Data Transmission Systems					●						●					●	●							●	●	●	■		
		14 Central Control and Display Equipment										●													●	●			●	■	
		15 Hardware Requirements																								●	●	●	●	●	■

Figure 1.1. Relationship between Chapters of this Handbook

select the chapter which has just been read, and then refer to the matrix for supplementary reading. The points on the matrix are those chapters which most closely relate to the principal chapter of interest.

NATURE OF THE PROBLEM

A freeway is designed for relatively high-speed vehicular flow between major trip-producing geographical areas. It is a multi-lane divided highway with grade-separated interchanges located at specific points to eliminate conflicting traffic. Ramps at these interchanges provide access to freeways from the standard highway network, and are designed to accelerate traffic to a suitable merging speed so that the mainline traffic flow is disturbed as little as possible.

Increasing population and increased vehicle ownership, have created new freeway design problems. The demand from an ever-increasing number of urban commuters has led to conditions that prevent many urban freeways from providing an acceptable level of service. The result has been congestion—the principal problem associated with urban freeways.

Congestion is expensive and socially undesirable in terms of driver delay, wasted gasoline, vehicular accidents, and atmospheric pollution. It is typically described in terms of both recurring and non-recurring problems (incidents).

Recurring congestion is caused by the overloading of a facility due to excessive demand or by geometric bottlenecks where sections of a facility have reduced capacity. After observation, time and location of this type of congestion can be predicted fairly accurately and managed effectively by controlling demand or increasing capacity.

Non-recurring congestion is caused by random events or incidents, that have the effect of reducing the capacity of a section of freeway. The most common types of incidents encountered include: traffic accidents, disabled vehicles, spilled loads, and adverse weather conditions. Since neither the location nor time of these random events is predictable, the resulting congestion cannot be dealt with by controlling demand or increasing capacity.

Identifying and Analyzing Problems	2. Data Collection 3. Delay Computation 4. Environmental Impacts 5. User Simulation Models
Procedures for Designing and Evaluating Solutions	6. Detection of Recurring Congestion 7. Ramp Control 8. Mainline and Corridor Control 9. Bus/Carpool Priority Control 10. Detection of Non-Recurring Congestion (Incidents) 11. Incident Detection Algorithms 12. Planning for Incident Response 13. Cost-Effectiveness Evaluation Techniques 14. Simulation Models in Design and Evaluation
Implementing Solutions	15. Financial Planning and Staged Construction 16. Interagency Cooperation

SECTION GUIDE

IDENTIFYING AND ANALYZING PROBLEMS

The next four chapters of this volume are concerned with the identification of problems and ways in which the extent of these problems can be quantified.

Chapter 2 describes the performance measures that can be collected for evaluating recurring congestion, and a technique that can be used to estimate the frequency of non-recurring congestion (incidents).

Chapter 3 focuses on a procedure for estimating delay due to incidents. The results are determined from the following input parameters: detection time, response time, clearance time, demand volumes, and capacity. Although this chapter is written with the analysis of different alternatives in mind, an analysis for detection, response, and clearance of existing delays created by incidents can be obtained by using average values for detection, response, and clearance time. Several examples have been included to help aid in the understanding of this material.

Chapter 4 discusses the noise and pollution aspects of freeway operation. It contains emission rate values for various vehicle operating modes. These can be applied to freeway operating conditions to determine the excess pollutants generated by congestion.

Chapter 5 describes the available freeway simulation models that can be used to identify and analyze conditions that cannot be directly measured. These include facilities not yet in operation, or those situations where evaluation of a future volume level is desired.

INTRODUCTION

Regardless of whether congestion is recurring or non-recurring, it is exhibited in terms of slow travel speeds, erratic speeds due to stop and go movements, increased and inconsistent traveltimes, increased accident potential, inefficient operation, and other undesirable characteristics.

Table 2.1 indicates data on the breakdown of the type of congestion in the Los Angeles area(1).

TABLE 2.1. TYPES OF CONGESTION

Type of Congestion	Typical Percentage Range
Recurring	43-57
Non-Recurring	
Incidents	30-47
Weekends	6-8
Holiday weekends	3-4
Others (e.g., special events)	1
Total Non-Recurring	40-60

The evaluation of congestion on existing freeway systems requires the identification of the causes and locations. For the planning of new systems, a traffic prediction followed by the use of a simulation model is necessary. Both traffic predictions and simulation models are subject to some uncertainty, they do, however, provide information on potential areas of congestion during the planning of new facilities.

MEASURES OF PERFORMANCE FOR RECURRING CONGESTION

The following sections from Everall(2) describe the concepts of the various measures that can be used to evaluate congestion.

Volume

One way of measuring system performance is to look at how the volume varies by time of day at a series of locations within the system. Such a plot is called a volume contour chart. The criterion for poor performance of the system might be when the volume exceeds a certain level, say 1,800 vehicles per lane per hour. This measure, however, does have the major deficiency of not indicating the location of congestion and traffic bottlenecks, since volumes of less than 1,800 vehicles per lane per hour can mean that either traffic is free flowing or the freeway is congested. It is recommended, therefore, that this measure not be used on its own, but in conjunction with another level of service measure.

Capacity

One important use of volume measurements is to derive capacity, defined to be the maximum traffic volume that can be handled by a particular roadway component under prevailing conditions. Using the analogy of momentum in physics, it is that product of the density and speed that maximizes the momentum of the traffic stream.

Demand

Another important measure of performance is traffic demand, also derived from volume measurements. Derivation of this measure is quite complex. The comparison of demand and capacity for the whole network under consideration can identify traffic problems. Figure 2.1 shows the relationship between demand, capacity, and congestion for a particular traffic bottleneck. It should be noted that congestion exists for a longer period than the time for which demand exceeds capacity. Thus, the important measures of performance here are the locations where demand exceeds capacity, the periods of time for which this occurs, and the periods of time for which congestion occurs.

Density

Density is another measurable quantity that can provide a good indication of how well the system is performing. It has been experimentally verified on California freeways(3) that densities less than 40 vehicles per lane per mile generally mean traffic flow is acceptable and speeds would be from about 40 miles per hour and above. (See figure 2.2 for the speed-density relationship.) Above 40 vehicles per lane per mile, a transition from heavy traffic to congestion occurs, until above 75 vehicles per lane per mile when there is severe congestion occurring upstream of a traffic bottleneck. Therefore, a suitable measure of freeway performance might be a density of 40 vehicles per lane per mile—densities above this would indicate stop-and-go traffic conditions. This value of density transition is dependent upon geometric factors, especially grades.

Occupancy

In like manner, occupancy can be a good measure of performance. Experience in Chicago(4) indicates that congestion on the freeway was impending when occupancy rose to between 15 and 25 percent as measured by ultrasonic detectors. (For loop detectors, the comparable impending congestion zone is 20 to 30 percent.) Athol(5) has noted that since traffic breakdowns may occur at less than maximum volume, the deviations from the linear relationship between volume and occupancy in stable flow conditions are perhaps the best sign of trouble. However, provided a sufficiently low threshold value of occupancy is chosen, occupancy can be a good parameter, indicating where bottlenecks and traffic congestion occur.

Speed

Another important parameter is speed, or its inverse equivalent, traveltime. Sampling of speeds over the roadways in the system can give a measure of where problems are apparent. For example, speeds can be measured at a number of points on a freeway, and one measure of performance might be the extent, both in time and space, of speeds below a certain level, say 40 miles per hour. However, a measure of this sort does require a large number of measurements that are not particularly easy nor efficient to make, since each point measurement really needs to be the average of a number of individual measurements (i.e., the measure of performance should be where average speeds drop below 40 miles per hour, rather than individual speeds below 40 miles per hour).

Again, if traveltime is being measured, then a suitable measure of performance obtained from individual representative runs might be on the freeway where traveltimes exceed 1.5 minutes per mile, or 3 minutes per mile on other arterial roads. Traveltimes can be expressed in terms of delay if some standard time is known for travel on the roads in the system in uncongested conditions. Thus, another measure of performance is the amount of delay per vehicle run through the system. Stopped time is also a measure of performance of the system, al-

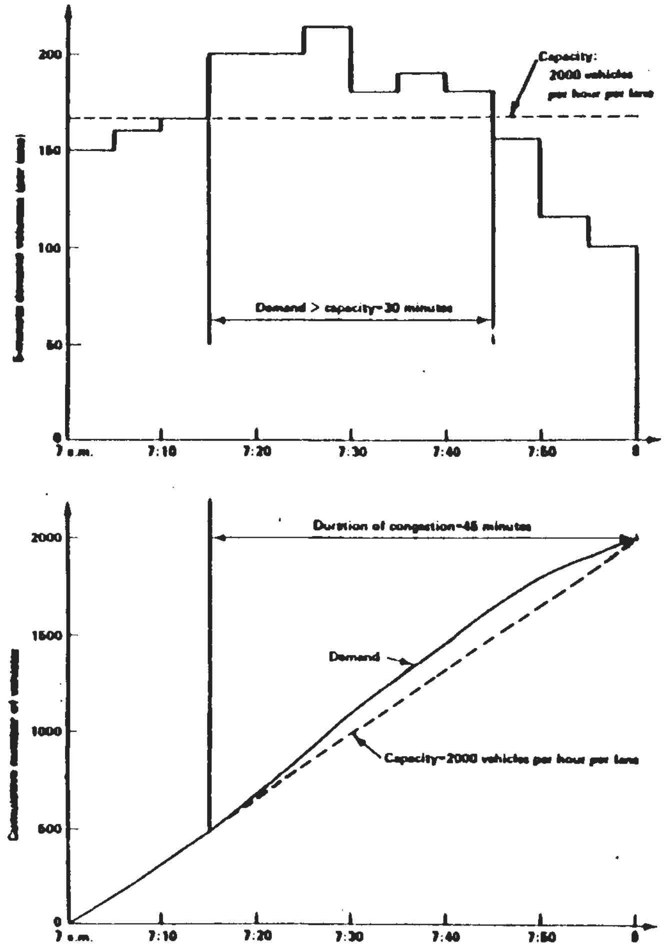


Figure 2.1. Relation Between Demand, Capacity, and Congestion

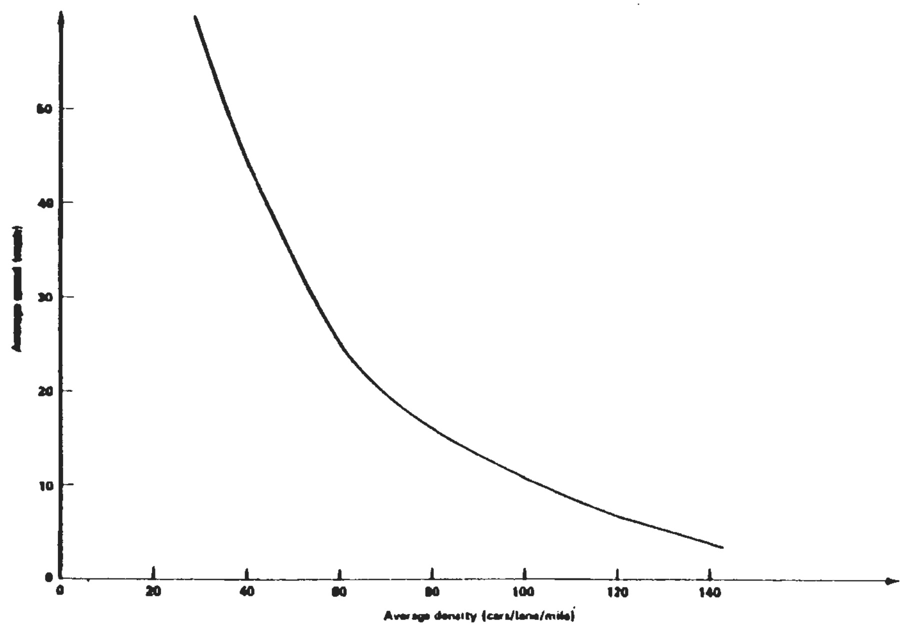


Figure 2.2. Speed-Density Relationship for Los Angeles Freeways'

though not a particularly good one, since it does not take account of any difference between free flow and congested (but not yet stop-and-go) conditions.

Throughput

From the point of view of the traffic engineer, perhaps the most important system parameter is the total amount of traffic carried by the network. This parameter has been termed "total travel"(6) and is measured in vehicle-miles per unit of time. This can either be expressed as a rate measured over a short period, often termed the "throughput," or as a total system parameter for a homogeneous period (say the morning peak period). The big advantage this parameter has over total volume of traffic in the system is that it takes account of the length of trips as well as the total number of vehicles. The most common derivation of this parameter is to:

- Divide the network into homogeneous sections (i.e., lengths in which there are no major traffic inputs or outputs).
- Measure or estimate the volume of traffic for the time period under consideration for each section.
- Sum the product of volume and section length for all sections.

Throughput can also be obtained from the product of the total number of vehicles in the system and a representative speed.

EVALUATION OF PERFORMANCE MEASURES

The measures of performance described so far have been in terms of the amount of traffic the freeway can handle, either on a time or distance basis. The individual motorist seldom appreciates the efficiency of a facility in terms of the volume accommodated, but is much more interested in the level of service provided —speed, comfort, safety, convenience, economy, etc. Level of service is one of the two major performance criteria suggested in the "Highway Capacity Manual"(7). The other performance criteria is capacity—but unfortunately it is not particularly useful in terms of making an inventory of freeway corridor problems since level of service is a qualitative, rather than a quantitative measure. Six levels of service—designated A through F—have been defined. It may be said that whenever a freeway (or other road) section is operating at levels of service D, E, and F, then congestion or unstable traffic conditions exist and there is a problem—but this still depends on the measurement of some other factor such as volume, density, or speed, from which a quantitative measure of performance can be derived.

Traveltime Analyses

From the point of view of the motorist, the important consideration is how much time is spent in the system. If this is summed for all motorists, then the important parameter obtained is the total traveltime within the system, measured in vehicle-hours per unit of time. The ideal way of obtaining this is to sum the traveltimes of all vehicles in the system, but this is not generally a practical method. A more common estimate of total traveltime is obtained by:

- Dividing the system into homogeneous sections
- Measuring the traffic volume on these sections
- Estimating or measuring representative traveltimes for each section
- Summing the product of volume and traveltime for all sections

A measure of total system delay is obtained by evaluating the total traveltime for the congested period with the equivalent summed product using representative uncongested traveltimes. The ratio of total travel to total traveltime is a measure of the average speed in the network.

The comparison of total travel and total traveltime is an excellent measure for evaluation of the merits of solutions to the freeway congestion problem. Figure 2.3 is a typical plot of total traveltime within a system against the total travel for a particular time period. Lines have been drawn on the diagram to represent values of average speed in the network. The triangles in the figure indicate values of total travel and traveltime for a network carrying traffic in the morning peak period. Depending on the amount of traffic, speeds vary between 17 and 25 mph (each point might represent data for one morning and measurements might be made over two weeks). If some form of solution such as ramp control or geo-

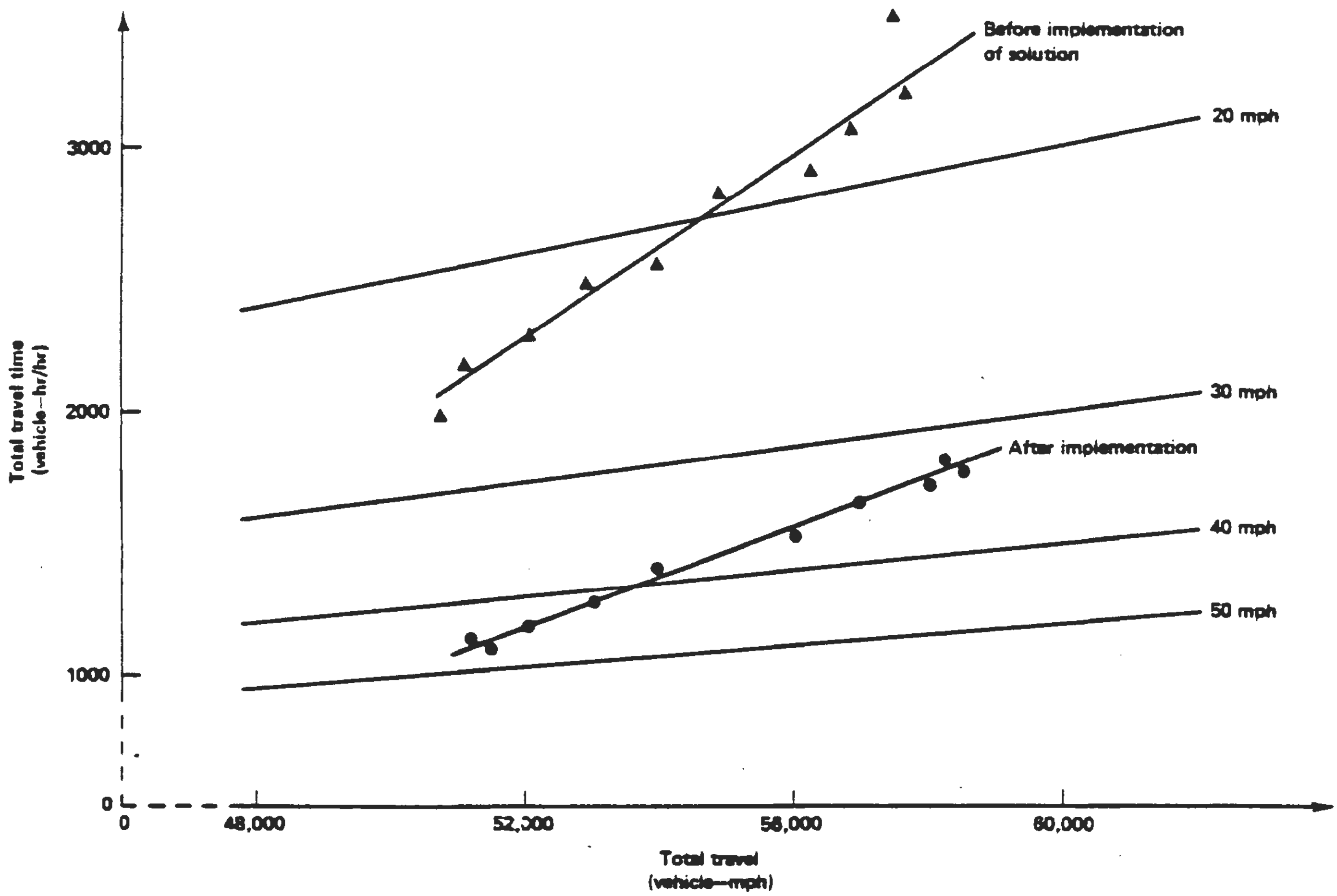


Figure 2.3. Comparison of Total Travel and Total Traveltime in a Network in the Morning Peak Period

metric improvement is applied which reduces congestion, then a different series of points will be obtained for the morning peak periods after implementation. Lines can be drawn through each set of points using linear regression analysis. Thus, if the amount of travel on the system remains the same, the ordinate difference between the two lines is an estimate of the total traveltime saved.

For example, in figure 2.3, with the before-and-after data shown, if the total travel on the system remains constant at 56,000 vehicle-miles per hour then implementation causes a reduction in traveltime of 1,400 vehicle-hours per hour, to which a dollar value can be attached.

Mode Shift Analysis

If a solution is being considered in which the modal split or vehicle occupancies may be changed, then it is better to measure total travel and total traveltime in terms of passenger-miles per hour and passenger-hours per mile rather than vehicle-miles per hour and vehicle-hours per hour. For example, if an alternative involves reserving one lane of freeway for buses only, then the correct measures of performance are in terms of passengers. The number of vehicle-hours that are spent in the system may go up but the total passenger-hours drop appreciably if there is a significant transfer of people to the buses.

Combined Performance Measures

So far, we have considered the two main measures of performance—maximum volume and maximum level of service—as independent parameters. A better measure of performance may be to optimize a function of both measures. The simplest measure of performance combining these parameters is the product of volume and speed. Since volume is equivalent to the product of density and speed, then this new measure is equivalent to the product of density and the square of the speed. Returning to the analogy with physics, this is then equivalent to kinetic energy, and the new parameter is referred to as the "kinetic energy" of the traffic stream. The aim of a project might, therefore, be to maximize the kinetic energy, which enables the concept of level of service to be optimized quantitatively. Drew and Keese(8) have compared kinetic energy with momentum (traffic volume). Figure 2.4 is such a comparison where each quantity has been normalized—speed by division by the free traffic speed u_f, momentum by division by the capacity q_m, and energy by division by $q_m u_f$. Based on fluid mechanics theory, the equations below have been developed for momentum q and energy q_u based on speed u and free speed u_f.

Normalized momentum:

$$\frac{q}{q_m} = 4 \frac{u}{u_f} - \frac{u^2}{u_f}$$

Normalized kinetic energy:

$$\frac{q_u}{q_m u_f} = 4 \frac{u^2}{u_f} - \frac{u}{u_f}^3$$

These equations give maximum momentum at an optimum speed of $1/2\ u_f$, and maximum kinetic energy at an optimum speed of $2/3\ u_f$. The volume at maximum kinetic energy may be termed the "optimum service volume," and the reduction in kinetic energy between the maximum and that at maximum momentum is analogous to internal friction in fluid mechanics. This energy-momentum analogy has been experimentally verified(8), and has been modified by Courage(9) to give separate curves for dry and wet weather.

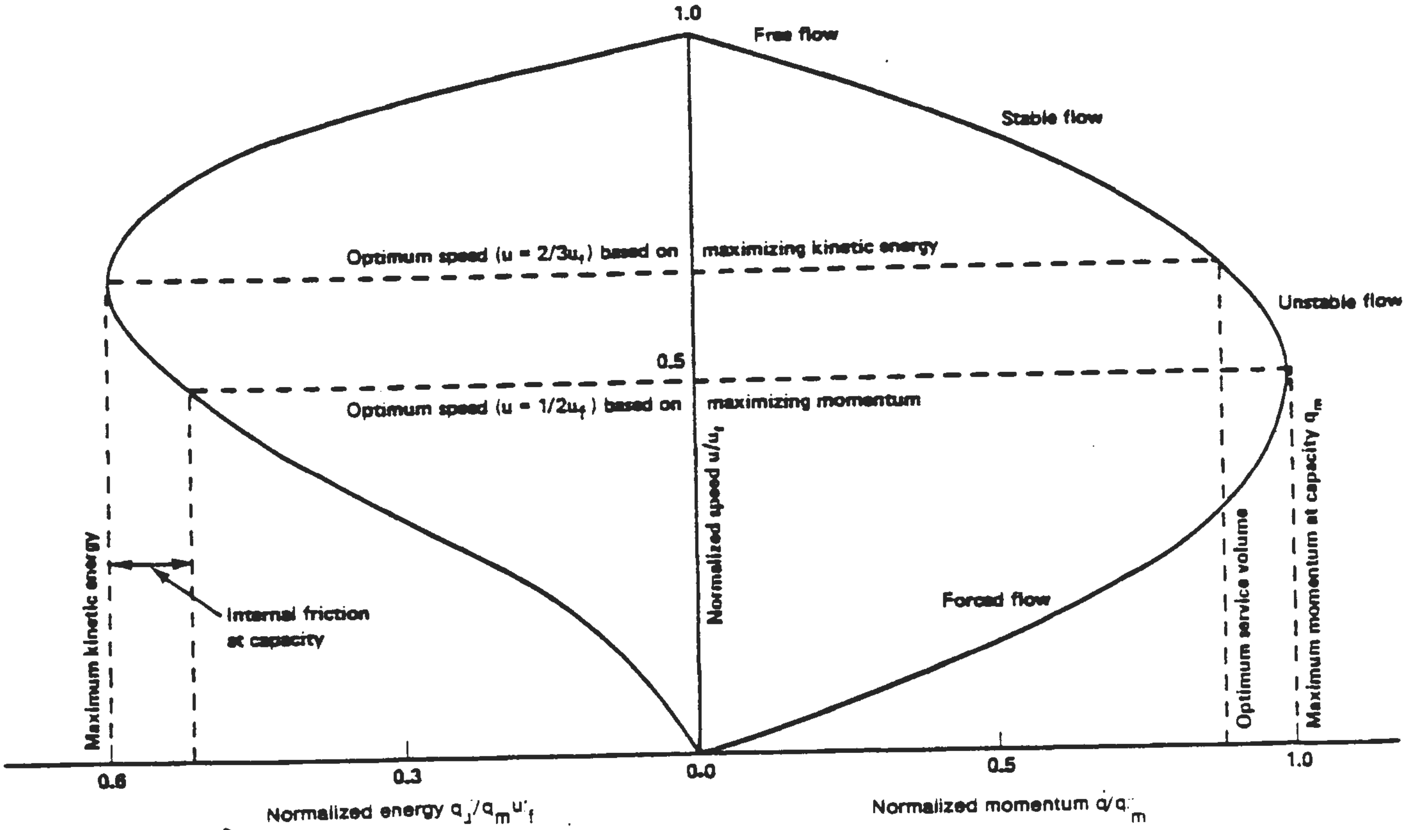


Figure 2.4. Comparison of Momentum and Energy as Measures of Performance

First-Cut Criteria

A simple first-cut measure of performance that can be used is the length of time congested conditions exist (i.e., the length of the peak period). Of course, congested conditions have to be defined—volumes above a certain level, speeds below a certain level, etc. Other possible definitions can be derived from the existence of queues on the ramps, on the freeway itself, or on other arterial streets. A more refined measure involves using queue lengths above a certain level or where the queues cause interference with other streams of traffic. Again, congestion can be defined in relation to the number of stops per vehicle-mile, and the length of time for which this parameter exceeds a certain level can be a measure of performance.

As outlined earlier, the reliability and predictability of peak-period traveltimes is important. This involves comparison of data obtained over a number of days, and such a measure of performance has rarely been used. For example, a suitable system measure for this would be the standard deviation of the total traveltime, measured over a three-week period. High values of this measure in relation to the mean total traveltime would indicate unpredictable service, caused either by large daily fluctuations in the total demand for travel or, more likely, a high frequency of incidents in the system. A simpler measure to obtain is the standard deviation of overall traveltime measured by making floating car runs at the same time each peak period. F and t statistics can be used to derive confidence limits for the variation of the means and standard deviations(10).

ACCIDENTS

The safety of a freeway section can be measured by analyzing the frequency of accidents resulting in fatalities, injuries, and/or property damage, as well as the number of deaths and injuries. These are then generally related to the amount of travel (millions of vehicle-miles) when highway lengths are being analyzed, and to the traffic volumes (millions of vehicles) for spot locations. Comparison with basic improvement criteria levels(11) will indicate whether these accident rates are such that remedies are needed. It is important to distinguish between fatal accidents and fatalities since accidents in certain locations may be relatively more severe, but in many cases calculation of the overall fatal and injury accident rates will serve to indicate where safety improvements may be necessary. If the number of conflicts is measured, these too should be expressed as a rate (conflicts per million vehicle-miles) and should be classified by type of conflict and type of roadway section.

When presenting accident data tables, engineers should be careful that the terms used are clear, as confusion often arises over fatal accidents that also involve nonfatal injuries. The following definitions from the FHWA(12) are commonly used:

- An <u>injury</u> is any bodily harm received by a person in a motor vehicle during a traffic accident.
- A <u>fatal injury</u> is any injury that results in death within one year of the accident.
- A <u>nonfatal injury</u> is any injury other than a fatal injury.
- A <u>fatal accident</u> is a motor vehicle accident resulting in one or more fatal injuries.
- A <u>nonfatal injury accident</u> is a motor vehicle accident that results in one or more injuries, none of which are fatal.
- A <u>fatality</u> is the death of any person who suffers a fatal injury.
- A <u>nonfatally injured person</u> is one who suffers a nonfatal injury in either a fatal accident or a nonfatal injury accident.

In order to evaluate an existing highway's accident record, it is convenient to estimate the existing accident data by comparison with national figures. Table 2.2 (12) shows the U.S. accident rates for 1980.

These figures are only suitable for looking at the accident record of major sections of freeway. In order to estimate how high or low accidents are at specific locations, a breakdown of accidents by location type must be used. Table 2.3 indicates likely accident rates by freeway location that the engineer can use as a reference point for planning purposes or responding to inquiries.

INCIDENT FREQUENCY

The initial step for the engineer in conducting a planning and trade-off analysis of incident management is to verify that such an analysis is warranted. To quantify the problem, the engineer first needs a reasonable estimate of the total number of incidents that can be expected to occur annually. This can sometimes be devel-

TABLE 2.2. INTERSTATE ACCIDENT RATES, 1980[1]

	Rural	Urban	Total
Highway Miles[2]	31,994	9,234	41,228
Annual Vehicle-Miles[2]	134,142	160,397	294,539
Daily Vehicle-Miles/Mile	11,456	47,460	19,520
Fatal Accidents			
- Number	1,913	2,017	3,930
- Rate[3]	1.43	1.26	1.33
Non-Fatal Injury Accidents[4]			
- Number	34,445	89,992	124,437
- Rate[3]	25.68	56.11	42.25
Fatalities			
- Number	2,300	2,225	4,525
- Rate[3]	1.71	1.39	1.54
Non-Fatally Injured Persons[4]			
- Number	55,911	135,882	191,793
- Rate[3]	41.68	84.72	65.12

1. U.S. estimates exclude the Commonwealth of Puerto Rico and the Territories of American Samoa, Guam, and Virgin Islands. Estimates for fatal accidents, fatalities, non-fatal injury accidents, and non-fatally injured persons are based on the partial data reported by states which are displayed in the following table together with totals reported by most states.

2. Mileage and travel data are from the Highway Performance Monitoring System (HPMS) for 1980. Federal-aid highway mileage is from HPMS universe data as of April 1, 1982, and vehicle-miles of travel are from the HPMS areawide summary tables as of April 1, 1982. Federal Highway Administration estimates were made for major highway categories where complete functional or Federal-aid system data were not reported.

3. Rates are per 100 million vehicle-miles.

4. Totals of non-fatal injury accidents and non-fatally injured persons were estimated by FHWA for Florida, Maryland, and Rhode Island, and for the non-state mileage of Missouri.

TABLE 2.3. FREEWAY ACCIDENT RATES BY LOCATION

Freeway Location	Number of Observations (Interchanges)	Average Number of Accidents Per Year	Accident Rate (Number of Accidents per Million Vehicles)
Full Cloverleaf	186	19.25	1.69
Partial Cloverleaf	191	5.33	0.94
3-Leaf or Trumpet	160	4.01	0.80
Full Diamond	681	4.12	1.02
Half Diamond	94	2.76	0.25
Full Slip Ramp Diamond	96	9.91	1.23

oped from local data that have been collected by the agencies or organizations involved in incident management. If reliable data are unavailable, however, the estimated incident rate presented in this chapter can be used by the engineer for planning purposes. Next, the number of delay-causing incidents must be estimated, since this subset of total incidents is of primary concern to the engineer. Once again, in the possible absence of empirical data, factors based on the detailed incident histories of certain surveillance and control projects have been provided. Finally, incident occurrence on individual segments of the freeway system can be projected such that the annual amount of delay may be determined.

Delay-Causing Incidents

For purposes of evaluating the potential effectiveness of certain incident management options, incidents that occur in a lane but are moved almost immediately to the shoulder are considered "shoulder incidents." Incidents on the shoulder do not usually create significant delay.

Given these basic assumptions regarding the definition of delay-causing incidents, the engineer can classify an incident according to: (1) its location on the roadway; (2) the degree of assistance necessary to effect its removal; (3) the type of incident; (4) the number of lanes affected; (5) the reason for a disablement; and (6) the nature of a mechanical problem. This classification method, which is presented graphically in figure 2.5 as an incident "tree," facilitates the identification of the percent and types of incidents for which remedial actions are effective. Each successive branch of the tree represents a finer disaggregation of incidents based on their delay-causing characteristics or the type of response that is required to effect removal. The tree was developed for two purposes: first, to organize an empirical incident data set into a form convenient for analysis; and second, to estimate the distribution of

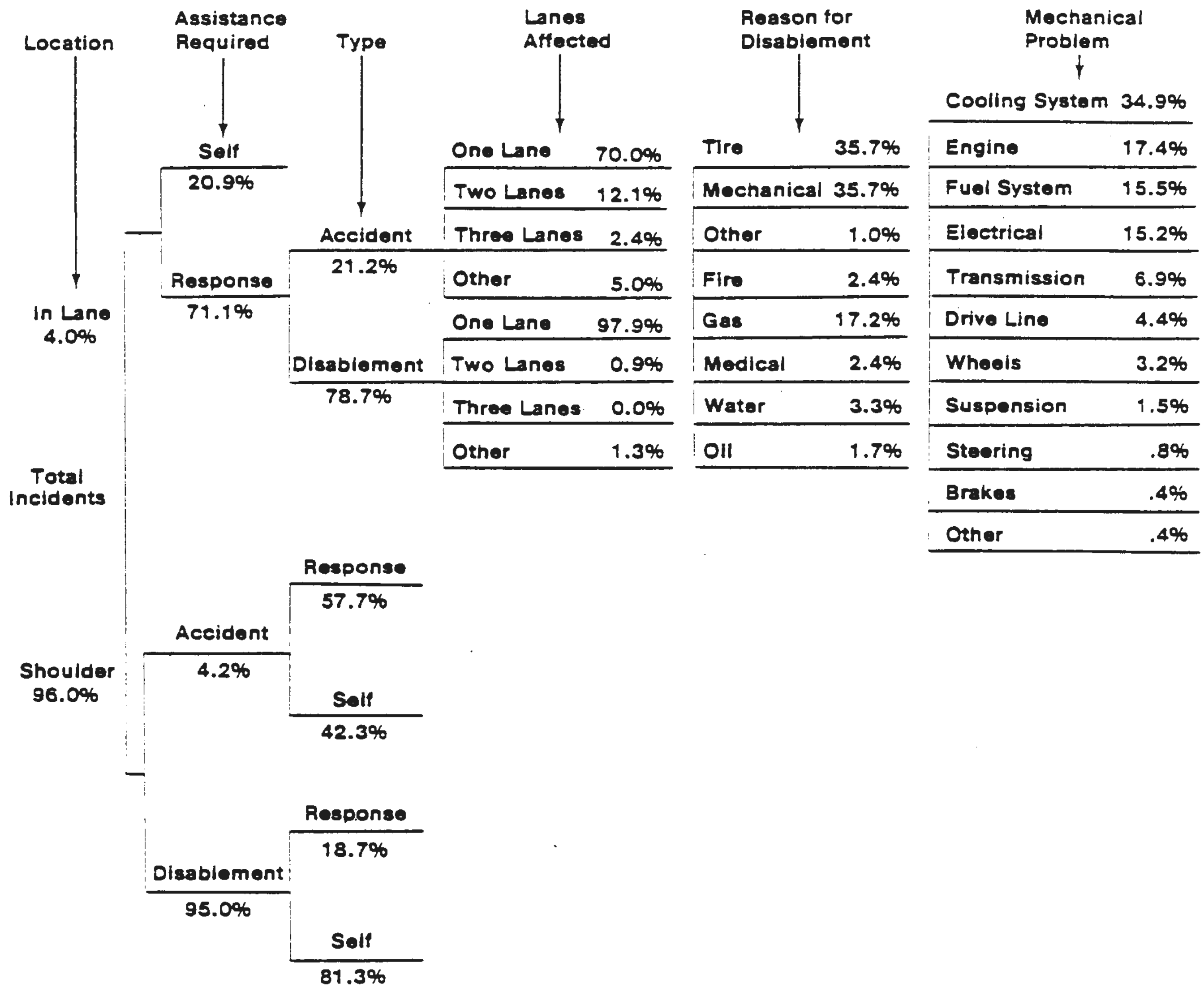


Figure 2.5. General Incident Tree

1 per 5000/miles

incident types in the absence of such detailed information for a specific facility.

Ideally, empirical incident data could be stratified according to the category branches presented in the incident tree. Data of sufficient detail to accomplish this arrangement, however, are unlikely to exist and are prohibitively expensive to collect. Therefore, for planning purposes, a typical distribution of incident types has been developed for use in evaluating incident management options. These factors, which are presented in figure 2.5, are based on incident data collected from a number of closed-circuit television surveillance projects throughout the country.

Freeway facilities without shoulders, such as bridges or tunnels, will experience fewer total incidents annually than those with shoulders and the distribution of types will change considerably. Incidents are estimated to occur at a rate equal to 39.3 percent of the rate for freeways with shoulders. Using the 200 incidents per million vehicle-miles (MVM) rate, recommended for planning purposes, this becomes 78.6 incidents per MVM. The distribution of incidents occurring on facilities without shoulders was calculated from data collected by the Detroit surveillance project and is shown in figure 2.6.

Incident Estimation by Segment

The number of incidents that can be expected to occur on a particular freeway system depends mainly on traffic volumes and roadway characteristics.

These two factors vary throughout a system according to time and/or location, so it is necessary to catalog them on a segment-by-segment basis in order to accurately estimate total annual delay.

Generally, the freeway system is easily divided into the highway segments between interchanges. If roadway charac-

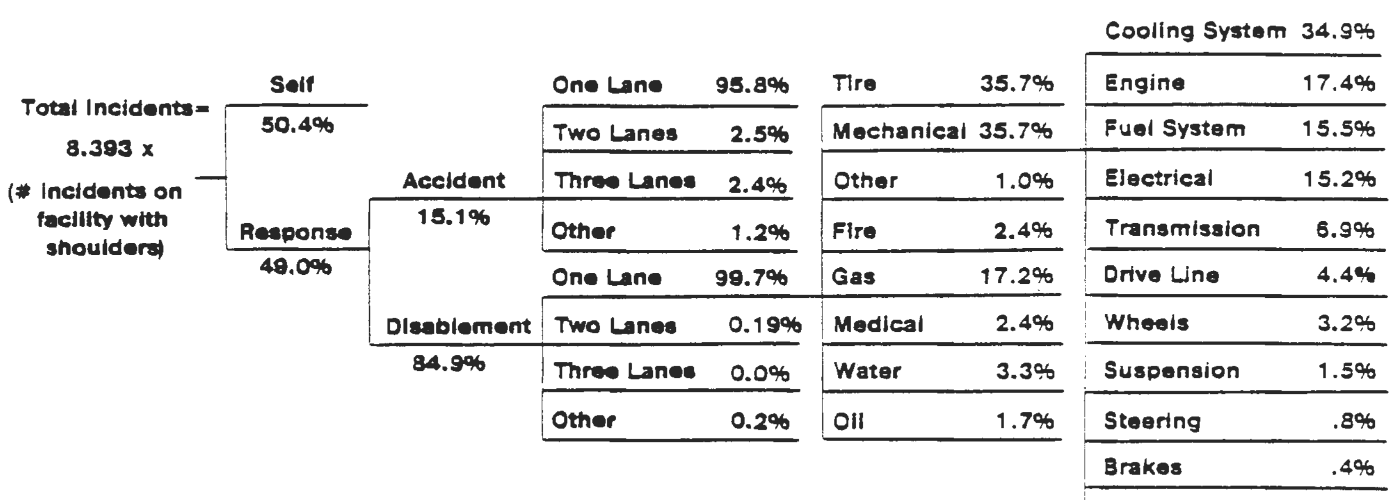


Figure 2.6. Incident Tree for Facilities Without Shoulders

teristics change over these distances (for example, the number of lanes or the existence of shoulders), then additional segmentation is recommended. Further division of highway segments may be necessary to account for changes in jurisdictional responsibility.

A standardized form should be used to record the volume and roadway characteristics of each segment of the freeway system. The roadway characteristics, including the number of lanes, length, and presence of shoulders, are listed along with information that identifies the segment and establishes its boundaries. In addition, the segment's vehicular volume by time of day and by direction should be measured or estimated.

An important criterion for continuing the analysis is whether traffic volumes are high enough to cause congestion and delay. Table 2.4 presents minimum flow rates beyond which congestion will occur in the event of an incident and, thus, enables the identification of those segments and time periods that are worthy of further consideration.

Once segment variations in traffic volumes and roadway characteristics have been catalogued, the annual number of incidents occurring on each segment can be estimated using the relationship illustrated in figure 2.7

TABLE 2.4. CONGESTION THRESHOLD VOLUMES

	Number of Lanes per Direction		
Incident Type	2	3	4
Accidents on Shoulder	3,000	4,600	6,300
In-Lane Incidents (all types)	1,300	2,700	4,300
Multi-Lane Accidents	All	1,200	2,600

(Values in the table represent volumes in vehicles per hour beyond which congestion will occur.)

Unless it has been determined empirically, the incident rate should be assumed as:

- 200 incidents per MVM for segments with shoulders

$$\text{Total Annual Incidents} = \frac{\begin{pmatrix}\text{Incident}\\ \text{Rate/MVM}\\ \text{(MVK)}\end{pmatrix} \times \begin{pmatrix}\text{Evaluation Period}\\ \text{Flow Rate}\\ \text{(veh/hr)}\end{pmatrix} \times \begin{pmatrix}\text{Length of}\\ \text{Evaluation Period}\\ \text{(hrs)}\end{pmatrix} \times \begin{pmatrix}\text{Evaluation Days}\end{pmatrix} \times \begin{pmatrix}\text{Length of Segment}\\ \text{in One Direction}\\ \text{(mi) (km)}\end{pmatrix}}{1{,}000{,}000}$$

Figure 2.7. Equation Used to Estimate Annual Incident Rate

- 78.6 incidents per MVM for segments without shoulders

Where the evaluation period flow rate is not known, an estimate can be obtained by using the average annual daily traffic (AADT) rate and procedures contained in the "Highway Capacity Manual." An evaluation period of two hours is usually satisfactory for peak-period analysis and 250 evaluation days per year used, unless local conditions dictate the use of another value.

Once the annual number of incidents occurring on a specific segment during a specific time period has been estimated, the total is distributed among the variety of incident types indicated in figure 2.5 or figure 2.6, the incident trees for facilities with and without shoulders, respectively. This information will provide a total picture of the number and type of incidents that can be expected, and enables an assessment to be made of the most appropriate incident management system.

Although an accurate estimate of the annual total number of incidents is valuable for identifying the magnitude of the freeway incident problem, it cannot, by itself, serve as a performance measure of a particular option or system. Effective incident management procedures are designed to reduce an incident's impact on traffic flow after it has occurred. Thus, a method for measuring and aggregating incident impacts is needed.

Delay computation methods that have been used to quantify these impacts are discussed in detail in chapter 5.

REFERENCES

1. Arceneaux, James L., and Mikhalkin, Basil. "Inventory of Freeway Geometric Bottleneck Congestion – Final Report." Freeway Operation Department, District 7, California Division of Highways. May 1973.

2. Everall, Paul F. Urban Freeway Surveillance and Control, The State of the Art.

3. Marino, R. "Inventory of Freeway Geometric Bottleneck Congestion – First Interim Report." California Division of Highways, Freeway Operation Department, District 7. February 1970.

4. May, A.D. "Experimentation with Manual and Automatic Ramp Control." Highway Research Record No. 59, 1964, pp. 9-38.

5. Athol, P. "Interdependence of Certain Operational Characteristics Within a Moving Traffic Stream." Highway Research Record No. 72, 1965, pp. 58-87.

6. Moskowitz, K. "Analysis and Projection of Research on Traffic Surveillance, Communication and Control." National Cooperative Highway Research Program Report 84, Washington, D.C., Highway Research Board, 1970.

7. "Highway Capacity Manual 1965." Highway Research Board Special Report 87, Washington, D.C., Highway Research Board, 1965.

8. Drew, D. R., and Keese, C. J. "Freeway Level of Service as Influenced by Volume and Capacity Characteristics." Highway Research Record No. 99, 1965, pp. 1-47.

9. Courage, K. G. "Some Electronic Measurements of Macroscopic Traffic Characteristics on a Multi-Lane Freeway." Highway Research Record No. 279, 1969, pp. 107-120.

10. Buhr, J. H., Whitson, R. H., Brewer, K. A., and Drew, D. R. "Traffic Characteristics for Implementation and Calibration of Freeway Merging Control Systems." Highway Research Record No. 279, 1969, pp. 87-106.

11. Roy Jurgenson and Associates, Westat Research Analysts, Inc. "Evaluation of Criteria for Safety Improvements on the Highway," 1966.

12. "Fatal and Injury Accident Rates on Federal-Aid and Other Highway Systems/1980," U.S. Department of Transportation. Federal Highway Administration. Office of Highway Planning, June 1982.

INTRODUCTION

This chapter discusses delay, the fundamental impact of freeway congestion and incidents. It presents a methodology for using delay as a performance measure to evaluate options.

Delay values can be converted, at least in an approximate manner, into other impact measures that can be aggregated for comprehensive evaluation. These include estimates of fuel consumption and air pollution caused by vehicles delayed by freeway incidents. Additionally, delay can be converted into user costs using estimating procedures similar to those often used in evaluating the costs of highway improvement projects or traffic operations alternatives.

The delay caused by recurring congestion can be determined with the aid of illustrations such as figure 2.1. In this illustration, delay is the area between the demand and capacity lines. For an actual data set, this area can be divided into a group of subareas for each time period with constant demand. The delay (areas) for each of these subareas can then be found by using basic algebraic equations.

DELAY DUE TO INCIDENTS

Any incident that blocks a freeway lane creates delay by reducing the number of vehicles that can pass that point in a given period of time. Even an accident removed to the roadway's shoulder will reduce capacity as motorists slow down to "rubberneck." In order to be quantified, delay must account for both the decrease in traffic volume and the duration of such a decrease. The delay caused by freeway incidents can be represented graphically in terms of traffic flow rates, as shown in figure 3.1. The horizontal axis is a time line indicating the occurrence of certain incident-related events and measuring the overall duration of incident-caused impacts on traffic flow. The vertical axis is the cumulative traffic volume, or, the sum of the vehicles passing any given point on the freeway in a defined time period.

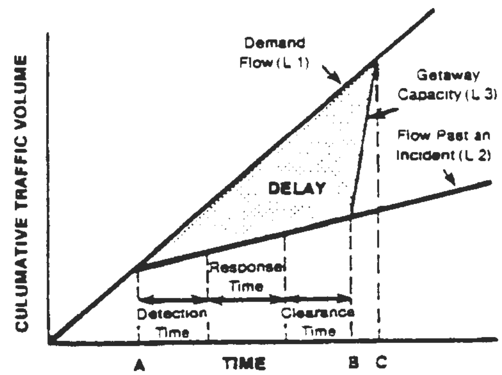


Figure 3.1.
Events Following an Incident

The total number of vehicles desiring to use the freeway, referred to as the demand flow rate, can be represented by line L1. When an incident occurs (time A), the traffic flow decreases below the demand flow due to a lane blockage. This reduced volume (shown as line L2 in figure 3.1) remains in effect until the incident is cleared from the freeway (time B). At that time, the queued traffic can begin flowing at a "getaway" rate approaching the freeway's capacity. At time C, the last vehicle in the queue reaches the normal flow speed, and traffic resumes flowing at the demand rate. The dotted area bounded by lines L1, L2, and L3 represents the total amount of delay (in vehicle-hours) created by the incident.

PROCEDURES FOR ESTIMATING TOTAL DELAY

A number of major factors affect the total amount of incident-caused delay as defined by figure 3.1. These include:

1. Freeway capacity
2. Demand flow (L1)
3. Reduced traffic flow rate (L2)
4. Getaway capacity (L3)
5. Duration of the incident

For the purpose of evaluation, freeway capacity is regarded as fixed. Unless it is possible for the demand flow to be diverted to other routes (with the possibility of other consequences arising), this may also be regarded as fixed in most freeway situations. Figure 3.1 shows the getaway capacity has a marginal effect on total delay once the incident has been cleared.

The two most important aspects in determining total delay are the reduced traffic flow rate and the duration time of the incident. These factors can be significantly affected by incident procedures.

The reduced traffic flow rate will be significantly affected by on-site incident management procedures. The duration of the incident is the result of the time taken to:

1. Detect the incident (referred to as detection time)
2. Respond to the incident once detected (referred to as response time)
3. Remove the incident from the freeway (referred to as clearance time)

In most situations, the greatest reduction in total delay can be achieved by decreasing incident duration. Of the components discussed above, detection time is frequently the longest, and incident detection procedures are discussed later. Factors that influence response time are discussed in chapter 12 of this volume. Clearance times are dependent on many factors and are difficult to shorten although on-site incident management procedures can also have an effect on this item. (This is discussed in volume 3.)

The remainder of this section is concerned with the procedures for estimating total delay once appropriate values, or a range of values, for the major variables have been determined. Delay can be estimated for a variety of management situations from a general condition, as shown in figure 3.2, as well as specific conditions, as appears in figure 3.3.

The general delay equation is designed to take account of the most complex conditions typically encountered. In a number of instances, a simplified version can be adopted. Such a simplified equation can be developed by considering four specific cases with reference to figure 3.3 and the general delay equation.

PLANNING THE APPLICATION OF DELAY EQUATIONS

The estimation of total delay is dependent on two types of variables: traffic flow rates and the duration of incident management activities. At least three, and up to five, flow rates (depending on delay conditions) must be known or estimated in order to calculate delay. Some of these can be easily measured in the field for particular freeway sections, but accurate values for the other rates are difficult to determine empirically without the benefit of a comprehensive study. Consequently, average flow rate values based on previous research efforts are provided in this section. The other determinant of delay, incident duration, defies generalization and must be estimated by planners familiar with the existing local incident management system. Incident duration can be viewed, however, in terms of its component incident management activities (i.e., see detection, response, and clearance times in figure 3.1), and consideration is given later to the factors that contribute to the time spent performing each of these actions. One of the primary goals of freeway incident management is to shorten incident duration by minimizing these individual component times.

Traffic Flow Rates

Traffic demand flow rates by time-of-day and freeway section should already be available from the cataloging exercise described in chapter 2. Capacity flow rates are well-documented by current highway research literature. These rates can be regarded as upper limits to the getaway rates maintained by queued vehicles passing the cleared incident site. For the purposes of this analysis and in the absence of reported research results, a getaway rate equal to the capacity flow rates is assumed. A value of 1,850 VLH for the capacity rate is recommended, which will conservatively estimate delay.

Values of the flow rate, S_3, for incidents that block one lane of freeway and for accidents on the freeway shoulder are shown in table 3.1.

To summarize, the traffic flow rates required for the calculation of total delay may be obtained as follows:

S_1 1850 VLH (table 3.1)

S_2 Initial Demand

S_3 See table 3.1

S_4 Depends on application

S_5 Typical range 15-50% of S_2

In all cases, local values of flow rates should be used if available.

TABLE 3.1. TYPICAL FLOW RATES FOR DELAY ESTIMATION

NUMBER OF LANES IN EACH DIRECTION	CAPACITY FLOW RATE = GET-AWAY FLOW RATE (VEHICLE/HR)	IN-LANE INCIDENTS (VEHICLE/HR)– ONE LANE BLOCKED	SHOULDER ACCIDENTS (VEHICLE/HR)
	S_1	S_{3a}	S_{3b}
2	3,700	1,300	3,000
3	5,550	2,700	4,600
4	7,400	4,300	6,300

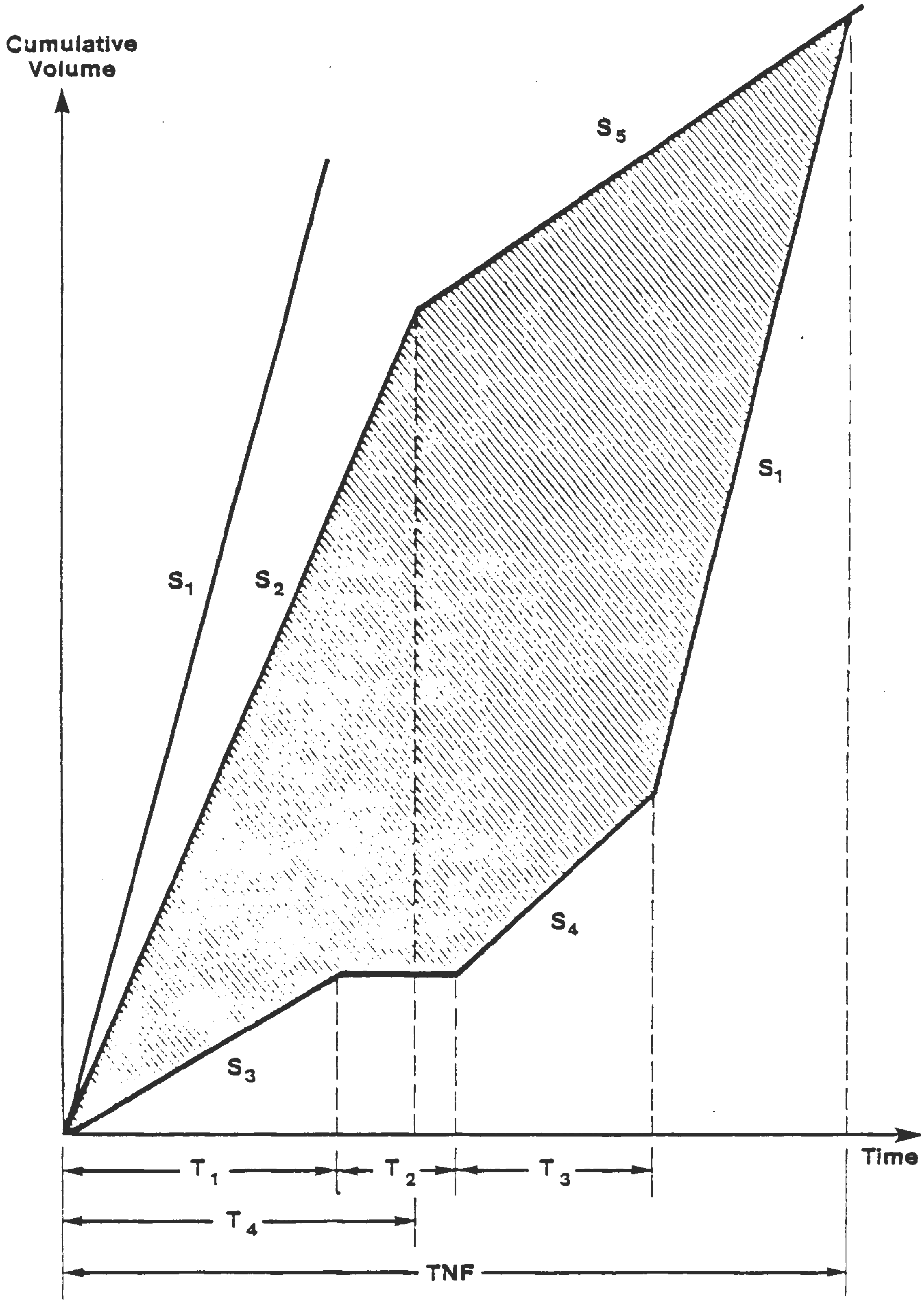


Figure 3.2. General Delay Condition

Figure 3.2, Continued

LEGEND:

S_1	-	capacity flow rate of the facility, vehicles/minute
S_2	-	initial demand flow rate, vehicles/minute
S_3	-	initial bottleneck flow rate, vehicles/minute
S_4	-	adjusted bottleneck flow rate, vehicles/minute
S_5	-	revised demand flow rate, vehicles/minute
T_1	-	incident duration until first change, minutes
T_2	-	duration of total closure, minutes
T_3	-	incident duration under adjusted flow, minutes
T_4	-	elapsed time under initial demand, minutes
D	-	total delay, vehicle-minutes
TNF	-	total elapsed time until normal flow resumed, minutes

Note: T_4 is independent of other times.

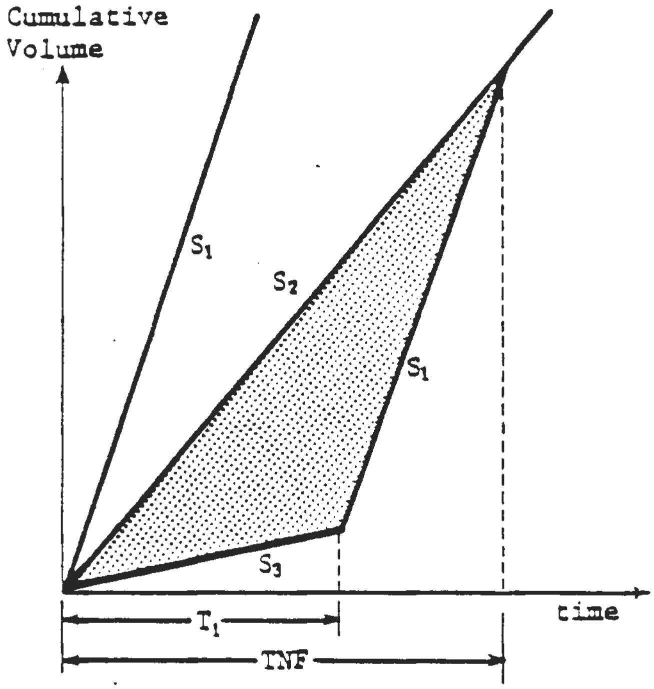
To compute delay under general conditions:

$$\begin{aligned} D = {} & T_1^2 (S_1-S_3)(S_5-S_3) + T_2^2 S_1 S_5 + T_3^2 (S_1-S_4)(S_5-S_4) - T_4^2 (S_1-S_2)(S_2-S_5) \\ & + 2T_1 T_2 S_1 (S_5-S_3) + 2T_1 T_3 (S_1-S_4)(S_5-S_3) \\ & + 2T_1 T_4 (S_1-S_3)(S_2-S_5) + 2T_2 T_3 S_5 (S_1-S_4) + 2T_2 T_4 S_1 (S_2-S_5) \\ & + 2T_3 T_4 (S_1-S_4)(S_2-S_5) \;/2 (S_1-S_5) \qquad (1) \end{aligned}$$

To compute time until normal flow resumes:

$$TNF = \frac{T_1 (S_1 - S_3) + T_2 S_1 + T_3 (S_1 - S_4) + T_4 (S_2 - S_5)}{(S_1 - S_5)} \qquad (2)$$

Figure 3.3. Delay Conditions



Condition 1: Simple Blockage

Condition 1 is known as a simple blockage. The number of vehicles that would have been processed if the incident had not occurred (the demand flow rate) is indicated by line S_2. The number that actually processed at the reduced flow rate is shown as line S_3. The duration of the incident, from the time of occurrence until clearance, is represented by the interval T_1. After the incident has been cleared, the queue of vehicles delayed by the incident will move past the site at a getaway rate of S_1, approaching the capacity flow rate of the facility. Traffic will continue to flow at this rate until all queued vehicles have been cleared.

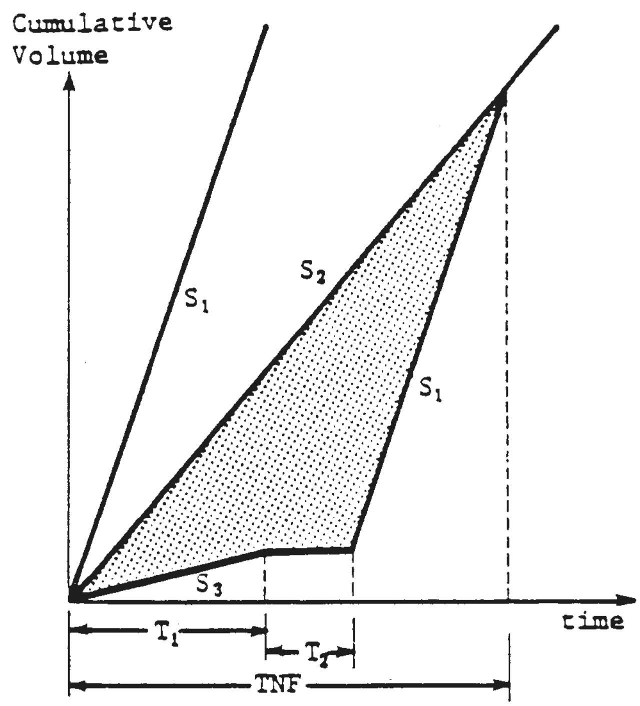
In this case T_2, T_3, and T_4 are zero; $S_2 = S_5$; and $S_3 = S_4$. The total delay is then given by:

$$D = \frac{T_1^2 (S_1 - S_3)(S_2 - S_3)}{2(S_1 - S_2)} \qquad (3)$$

The total elapsed time until normal flow resumes is calculated by:

$$TNF = \frac{T_1 (S_1 - S_3)}{(S_1 - S_2)} \qquad (4)$$

Figure 3.3, Continued



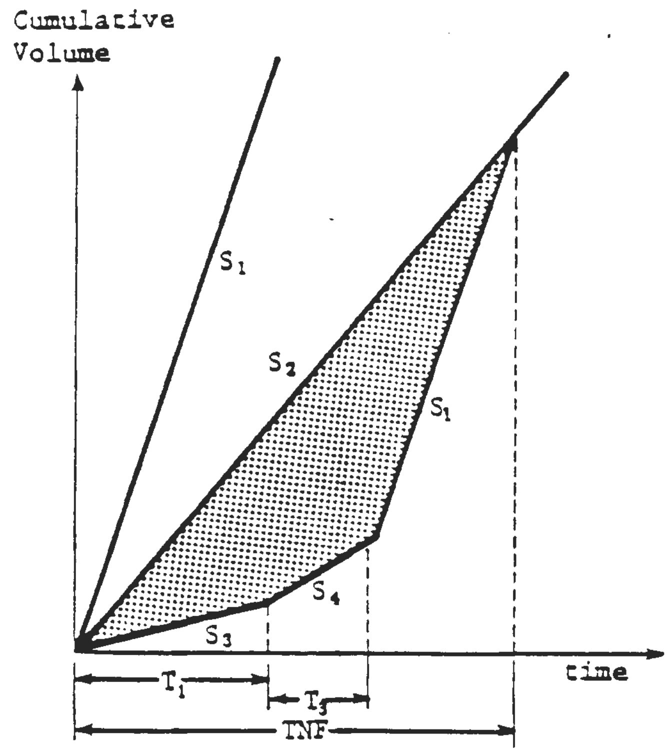
Condition 2: Short-Term Closure

In this condition the freeway is entirely closed for a period T_2 in order to remove the incident. Once cleared, flow returns to the getaway value S_1. In this case, T_3 and T_4 are zero, and $S_2 = S_5$, and $S_4 = S_1$. Therefore:

$$D = \frac{T_1^2 (S_1 - S_3)(S_2 - S_3) + T_2^2 S_1 S_2 + 2T_1 T_2 S_1 (S_2 - S_3)}{2(S_1 - S_2)} \tag{5}$$

$$TNF = \frac{T_1 (S_1 - S_3) + T_2 S_1}{(S_1 - S_2)} \tag{6}$$

Figure 3.3, Continued



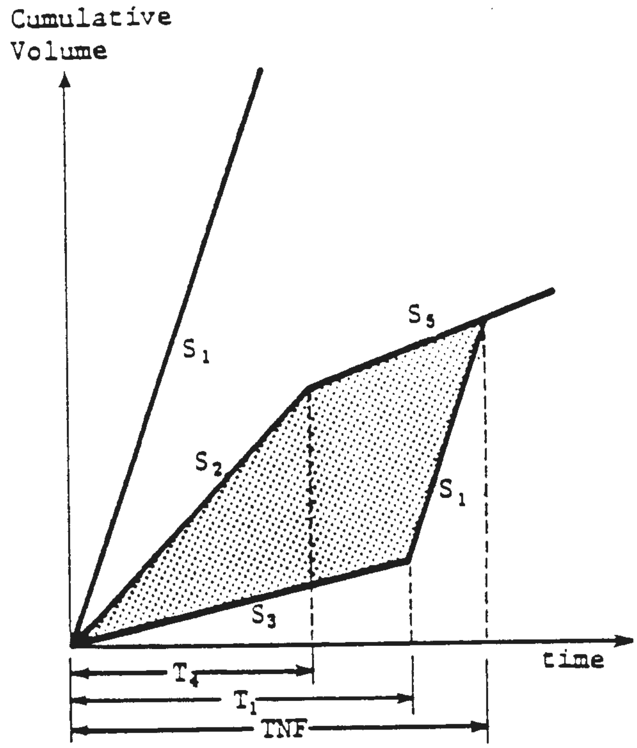
Condition 3: Adjusted Bottleneck

In this case the capacity available is increased prior to total clearance of the incident. For example, two lanes may have been blocked initially but one is cleared prior to total removal of the incident. In these circumstances T_2 and T_4 are zero and $S_2 = S_5$, hence:

$$D = \frac{T_1^2 (S_1 - S_3)(S_2 - S_3) + T_3^2 (S_1 - S_4)(S_2 - S_4) + 2T_1 T_3 (S_1 - S_4)(S_2 - S_4)}{2(S_1 - S_2)} \qquad (7)$$

$$TNF = \frac{T_1 (S_1 - S_3) + T_3 (S_1 - S_4)}{(S_1 - S_2)} \qquad (8)$$

Figure 3.3, Continued



Condition 4: Revised Demand

This case illustrates how total delay decreases when the demand rate is reduced during the incident by planned or unplanned diversions of upstream traffic or by a decrease in demand at the end of a peak period. In this case, $T_2 = 0$ and $S_3 = S_4$, and the equations become:

$$D = \frac{T_1^2 (S_1 - S_3)(S_5 - S_3) - T_4^2 (S_1 - S_2)(S_2 - S_5) + 2T_1 T_4 (S_1 - S_3)(S_2 - S_5)}{2(S_1 - S_5)} \tag{9}$$

$$TNF = \frac{T_1 (S_1 - S_3) + T_4 (S_2 - S_5)}{(S_1 - S_5)} \tag{10}$$

Tabulated Values

Using equations (1) and (2), previously computer-generated tables of delay and the time until normal flow is resumed have been produced and are presented in volume 6 of Alternative Surveillance Concepts and Methods for Freeway Incident Management(1). These tables apply only to Condition 1: Simple Blockage, and Condition 4: Revised Demand. They cover a wide range of roadway and flow conditions, because of space limitations, they have not been reproduced in this handbook.

For most applications, equations (1) to (10) can be solved with a calculator. If an agency desires to prepare a simple computer program to compile a set of tables for appropriate ranges to suit local circumstances, reference should be made to appendix 2 of volume 2 of reference 1, where a number of special conditions and considerations are noted.

EXAMPLES

This portion of the chapter uses examples to explain how to perform the necessary analysis for determining the number of vehicle-hours of delay caused by incidents on a freeway network; examine this delay for possible reductions; and choose and evaluate options for implementation that will reduce such delay. Material is also included on options that cannot be evaluated easily in quantitative terms. This material is derived principally from a report by Peat, Marwick, Mitchell & Co. The original work entitled, Alternative Surveillance Concepts and Methods for Freeway Incident Management(1) was prepared for the Federal Highway Administration and contains six volumes, to which the engineer is referred for further information.

Delay Computation Worksheet

In order to organize the results of the evaluation process, it is recommended that a delay computation worksheet (DCW) be used. A sample DCW is provided in figure 3.4.

The delay computation worksheet is divided into five sections, and the purpose of each is described briefly in the following paragraphs.

The first section is used to identify the segment being evaluated and to assess detection capabilities. The major product of this section is an estimate of how long it will take the existing system to detect an incident on this segment.

The second section, Weighted Service Time, is an optional section used only if the existing detection systems or environment surrounding the segment offer a particular service that will affect the average detection, response, or clearance time for any of the existing services. For example, if certain police vehicles have push bumpers, the response time is zero for pushable incidents, since a policeman can clear the incident upon detection by pushing it out of the traffic lane. Thus, for this example, the entries in this section will produce a response time for police-detected incidents that takes into account the zero response time for pushable incidents.

The third section, Delay Computation Variables, defines the volume characteristics of the segment under evaluation. These characteristics indicate the capacity of the segment, the demand volume on the segment both during and after the evaluation period, the flow rate around an in-lane incident and around a shoulder accident, and the duration of the initial demand flow rate. All of this information

DELAY COMPUTATION WORKSHEET

1

Segment ______________ Time Period ________________________ Lanes ______ Shoulders Y/N

System/Option ____

P_1 (__________) = __________ %, P_2 (__________) = __________ %, A_2 = __________ Min.

Headway = __________; R_1 = 1/__________ Expected Detection Time = __________ Min.

2 WEIGHTED SERVICE TIME (Minutes)

TIME	% of Incid.	Nom. Time.	Product	% of Incid.	Nom. Time	Product	% of Incid.	Nom. Time	Product	TOTAL

3 DELAY COMPUTATION VARIABLES

S_1 (Veh/Hr)	S_2 (Veh/Hr)	S_5 (Veh/Hr)	S_3 (Veh/Hr) Shldr	S_3 (Veh/Hr) Lane	T_4 (Min.)

4 INCIDENT SERVICE TIMES FOR EACH DETECTION METHOD (Minutes)

SHOULDER ACCIDENTS – No. __________

	Detection	+ Response	+ Clearance	= TOTAL
T_1: ALL/POLICE				
T_1: OTHER				
T_1: SPECIAL __________				

IN-LANE INCIDENTS – No. __________

	Detection	+ Response	+ Clearance	= TOTAL
T_1: ALL/POLICE				
T_1: OTHER				
T_1: SPECIAL __________				

5 TOTAL DELAY FOR EACH DETECTION METHOD (Vehicle Hours)

	SHOULDER ACCIDENTS			IN-LANE INCIDENTS		
	All/Police	Other	Special	All/Police	Other	Special
DELAY/INCIDENT						
NO. OF INCIDENTS/YEAR						
DELAY/YEAR						

BASE DELAY: ________________ – OPTION DELAY: ________________ = ________________ A.V.H.

COST OF THIS OPTION ____

NOTES: ____

Figure 3.4. Blank Delay Computation Worksheet

is used in the fifth section of the worksheet to determine the estimated delay.

The fourth section, Incident Service Times for Each Detection Method, is used to determine the total number of incidents occurring on the segment under evaluation and an average duration time for these incidents. As in-lane incidents create greater delay than accidents on the shoulder, the two categories are evaluated separately.

The fifth section, Total Delay for Each Detection Method, brings together all of the factors included in the previous sections to determine the number of vehicle-hours of delay resulting from the estimated number of incidents and other characteristics of the segment. This section will produce one of two results, depending upon the evaluation circumstances. If the base case or existing system is being evaluated, the result of this section will be a value indicating the number of vehicle-hours of delay currently being incurred on the segment. If an option is being evaluated, the result will be the number of vehicle-hours of improvement that are likely to result from implementation of the option.

Illustrative Example: 10-Mile Freeway Segment—Delay Evaluation

Defining the Study Area and Analysis Segments — The first step in the evaluation process is to prepare a thorough description of the study area and to check that the area initially defined is, in fact, adequate for the purpose in mind. Among the factors that should be given particular consideration are:

- Administrative boundaries
- Major highways and type
- Incidence of congestion
- Definition of affected organizations and their geographic areas of responsibility (police, fire, rescue, ambulance, public works departments, etc.)

Within the study area, each facility selected for evaluation must be divided into a series of segments for analysis purposes. A segment can be defined by considering the following factors:

- Major political boundaries (state lines, county boundaries)
- Major service area changes (service areas for police, fire, paramedic, wrecker, etc.)
- Major roadway characteristics (number of lanes, presence of adequate shoulders, frontage roads)
- Major traffic characteristics (primarily volume)

Before evaluation can be undertaken, the final decisions that must be made relate to the selection of time periods to be included. Time periods can be selected based on planning judgment, or they can be related to specific physical and traffic characteristics, such as:

- Traffic volume fluctuations by hour and day
- Patrol (surveillance) level differences
- Periods of extraordinary incident frequency, such as after dark, in bad weather, during holiday periods, etc.

At a minimum, the time periods must reflect major changes in volume levels and patrol frequencies. Delay impacts might be significant only during certain periods, depending on traffic and roadway characteristics, and the evaluation could therefore, be limited to these periods. Alternatively, a preliminary analysis might be undertaken for all time periods and then dropped for those in which no improvements are necessary.

An important criterion in selecting time periods for analysis is whether traffic volumes are high enough to cause congestion and delay. A simple table showing volume levels beyond which congestion will occur is given in the "Highway Capacity Manual."

For the illustrative example, the option of a police service patrol has been used; however, the calculations would be very similar for a dedicated freeway patrol. The general and specific characteristics of the freeway segment are shown in table 3.2, and it is assumed to have the following features:

- It is 10 miles long with 10-foot shoulders.
- Medians are 30 to 60 feet of grass, allowing incident detection across them.
- The segment is three lanes in each direction.
- Peak-hour directional volume is 5,000 vehicles per hour.
- Peak-hour durations are two hours each.
- Volumes are 2,500 vehicles per hour during peak periods in non-peak direction of flow.
- The police patrol operates on 90-minute headways.
- The police cruiser is equipped with push bumpers.
- Peak volumes are nearly identical (see table 3.2), so computationally, the number of incidents can be doubled to obtain total incidents.
- The period of evaluation is assumed to be 250 days per year.

Completion of the DCW

The completed DCW for the initial conditions on the 10-mile segment is shown in figure 3.5. Each entry on the form has an identifying number in a circle that corresponds with an explanation of that entry in the following text.

Section 1 of the DCW (Base Case)

1 List number identifying segment.

2 Time periods of analysis; as volumes are similar, a.m. peak (EB) and p.m. peak (WB) are analyzed together in the example (see table 3.2)

3 List travel lanes in each direction

4 Yes means shoulders 10 feet or more wide.
No means shoulders less than 10 feet wide.

5 The existing situation, or base case, is being evaluated to determine the magnitude of the problem.

6 7 Defines the first mode in the calculation of detection time. In the example, P_1 has been defined as the State Police; and previous records show, in the study area, that the State Police are the first to detect 6 out of every 10 incidents (i.e., 60 percent).

8 9 Defines the second mode in the calculation of detection time. In the example, this has been assumed to be all other existing detection methods (e.g., passing motorists), and hence 4 out of every 10 incidents are detected this way (40 percent). Note $P_1 + P_2$ must equal 100 percent.

GENERAL CHARACTERISTICS OF SEGMENT

Route: I - XX From: milepost 70 To: milepost 80

Lanes: 3+3 Length: 10 mi. Shoulders: Yes Frontage Roads: No

Interchange Spacing: 1.5 mi. ADT: unknown Other:

Police Agency:

Freeway Itself	Adjacent Streets
State police	City police

Service Patrol: None

Traffic Engineering/Operations:

Freeway Itself	Adjacent Streets
DOT	City

Fire/Rescue Services: County volunteer

Wrecker Services: Rotation list

Construction/Maintenance: DOT

Other Agencies:

Call Boxes: (Y/N) N Spacing: Monitoring: Maintenance:

Other Resources:

Services Provided – Police:

Gas (Y/N) N Change Tires (Y/N) N Push Bumpers (Y/N) Y Other

Services Provided Service Patrol:

Push Bumpers (Y/N) Minor Engine Repairs (Y/N) Other

SPECIFIC CHARACTERISTICS OF SEGMENT

	Time Period 1	2	3	4	5	6
	7-9	9-5	5-7	7-7		
Volume (veh/hr):						
Direction E	5,000	2500	3,000	4,000		
Direction W	2,950	2,500	5,000	4,000		
Post Peak	1,500					
Police Patrol Headway (min.):	90	90	90	90		
Service Patrol Headway (min.):	—	—	—	—		
Average Operating Speed (mph):	45	55	45	55		
Average Response Time (min.):						
Wrecker	10	10	10	15		
Ambulance	10	10	10	10		
Fire/Rescue	5	5	5	5		
Detection Probabilities:						
Police	60%	60%	60%	60%		
Service/Other	40%	40%	40%	40%		
Special						
Clearance Time: (min.)	20	20	20	20		

Notes:

Table 3.2. General and Specific Characteristics of 10-Mile Segment

DELAY COMPUTATION WORKSHEET

1
Segment (1) 92-8 Time Period AM (EB) (2) PM (WB) Lanes (3) 3 Shoulders (4) Y/N

System/Option Base Case (5)

P_1 (Police (6)) = 60 (7) %, P_2 (Other (8)) = 40 (9) %, A_2 = 10 (10) Min.

Headway = 90 (11); R_1 = 1/45 (12) Expected Detection Time = 27 (13) Min.

2 WEIGHTED SERVICE TIME (Minutes)

TIME (14)	% of Incid. (15)	Nom. Time (16)	Product (17)	% of Incid. (18)	Nom. Time (19)	Product (20)	% of Incid.	Nom. Time (21)	Product	TOTAL (22)
Resp: Police	56	0	0	44	10	4.4				4.4
Resp.-Other	56	5	2.8	44	10	4.4				7.2

3 DELAY COMPUTATION VARIABLES

S_1 (Veh/Hr)	S_2 (Veh/Hr)	S_5 (Veh/Hr)	S_3 (Veh/Hr) Shldr (26) Lane	T_4 (Min.)
5550 (23)	5000 (24)	2500 (25)	4600 / 2700	60 (27)

4 INCIDENT SERVICE TIMES FOR EACH DETECTION METHOD (Minutes)

SHOULDER ACCIDENTS — No. 233 (28)

(30)

(29)	Detection	+ Response	+ Clearance	= TOTAL
T_1: ALL/~~POLICE~~	27	10	20	57
T_1: OTHER				
T_1: SPECIAL				

IN-LANE INCIDENTS — No. 267 (31)

(32)

	Detection	Response	Clearance	TOTAL
T_1: ~~ALL~~/POLICE	27	4	20	51
T_1: OTHER	27	7	20	54
T_1: SPECIAL				

5 TOTAL DELAY FOR EACH DETECTION METHOD (Vehicle-Hours)

(33)	SHOULDER ACCIDENTS All/~~Police~~	Other	Special	IN-LANE INCIDENTS ~~All~~/Police	Other	Special
DELAY/INCIDENT	219			1692 (34)	1901	
NO. OF INCIDENTS/YEAR (35)	233			160	107	
DELAY/YEAR (36)	51027			270,720	142,707	

(37) BASE DELAY: 514,454 — OPTION DELAY (38) N. A. = N. A. A.V.H. (39)

COST OF THIS OPTION N. A. (40)

NOTES: (41)

Figure 3.5. DCW for 10-Mile Segment Base Case

(10) This is the average reporting time for all other detection methods (A_1 is zero since the police have zero reporting time). In the example it has been estimated by examining dispatcher logs that 50 percent of the incidents reported by passing motorists are reported by telephone, and the other 50 percent are reported by motorists flagging an officer on the roadway. From empirical field work, it was determined for this example that it would take the passing motorists 15 minutes to reach a phone and make a call and five minutes for motorists to flag a patrolman. An average of these values results in the 10-minute entry.

(11) The headway between police patrol vehicles (which is the total time to complete one circuit of the patrol beat) is shown as 90 minutes for this example in table 3.2.

(12) R is the effective patrol frequency and is defined as 1/headway. If, as in the example, there is good visibility across the median and an incident can be detected by a police officer traveling in either direction, the effective patrol frequency is twice as high and becomes 1/(½ headway). Thus, the entry in the example is 1/45.

(13) The expected detection time is the principal output from Section 1 of the DCW. For the two-mode situation in the example, it can be computed by using the equation:

$$EX = P_i/R_i = 0.6/(1/45) = 27 \text{ minutes}$$

For convenience, figure 3.6 may also be used by entering the effective headway and reading off the expected detection time corresponding to the appropriate P_i value.

Section 2 of the DCW (Base Case)

Before proceeding further, a calculation or determination must be made regarding incident distribution or the values applied to the incident in Figure 2.5, General Incident Tree. That is, either site-specific values must be determined for all or part of the items indicated on figure 2.5 or the indicated default values must be used. Police logs will usually contain a distribution of response-required incidents. However, information usually does not exist regarding in-lane versus shoulder splits (although this can sometimes be inferred). No information regarding the response-self splits will be found in the logs because "self" incidents solve themselves, and lane information is likely to be reported for only large incidents. The best information that can be expected from police logs is a fairly reliable distribution of the cause of in-lane, one-lane response-required disablements and total accident information. Therefore, to begin this portion of the analysis, the first step is to either use the indicated default values or refine the values. For this illustrative example, it is assumed that the indicated percentages on the incident tree reflect the correct distribution for the 10-mile segment.

Section 2 is completed only if a particular service exists on the primary patrol vehicle (P_1) or exists in the operating environment (e.g., a privately owned and financed tow truck patrol). A listing of these particular services includes push bumpers, gasoline-dispensing equipment, water, and oil carried on the patrol vehicle; a fully equipped service patrol; availability of accident investigation sites; and the existence of a fast vehicle-removal law.

(14) This entry identifies the mode in which the response time is affected by a particular piece of equipment or operating environment. In the

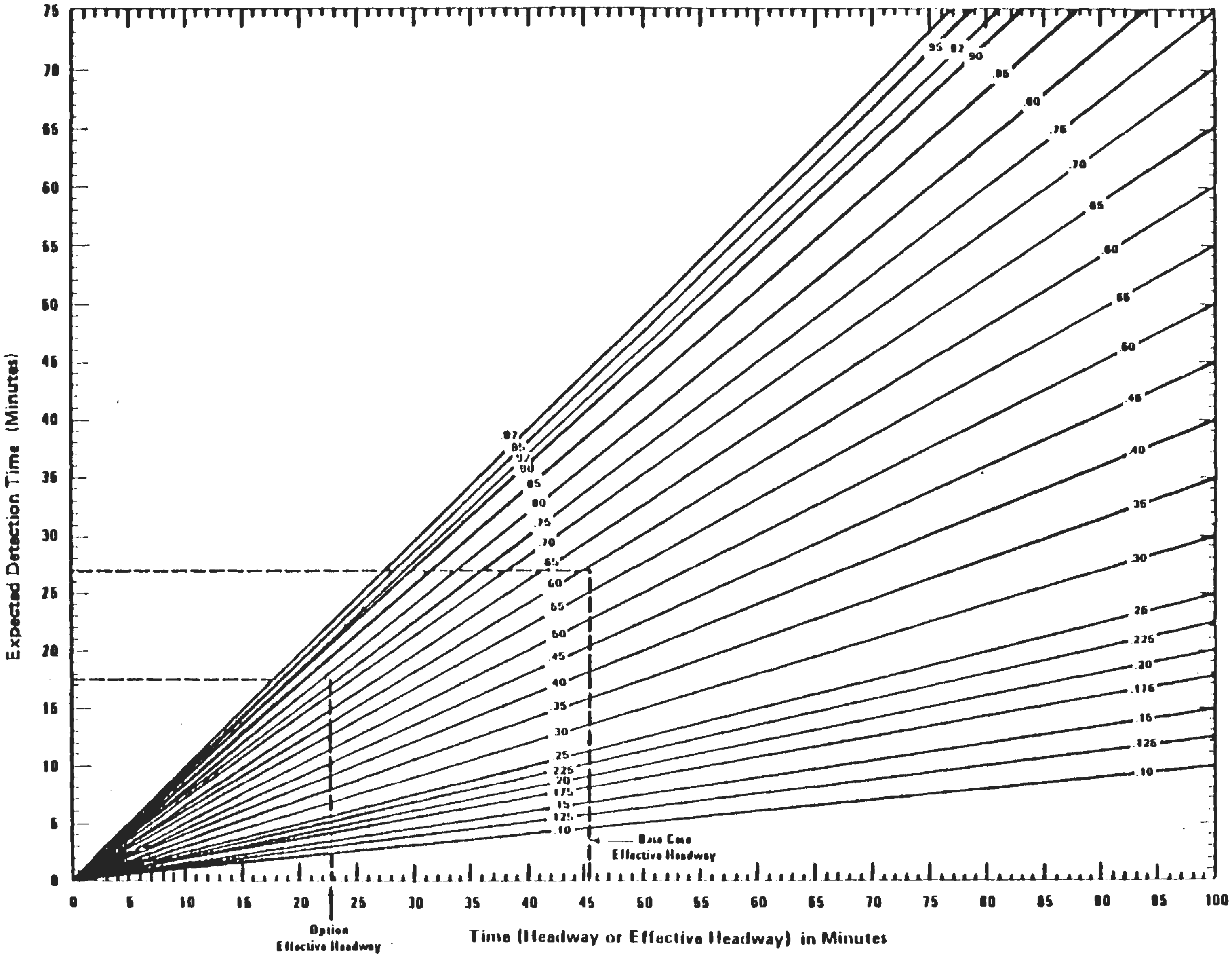


Figure 3.6. Expected Detection Time Computation

example it was assumed that the police patrol vehicles were fitted with push bumpers. Thus, a police patrol detecting a "pushable" incident has a zero response time.

15 This entry contains the percentage of incidents affected by the conditions identified in entry 14. This can be determined from the information on incident patterns in figure 2.5 if local data are not available. Using this pattern and, for the purposes of the example, considering incidents that only affect one lane, the proportion of pushable incidents may be determined as follows:

Proportion of incidents affecting one lane

Accident:	0.213 x 0.799 =	0.170
Disablement:	0.787 x 0.979 =	0.770
Total		0.940

Proportion of disablements that can be pushed

Gas (17.2%)
Water (3.3%)
Oil (1.7%)
Other (1.6%)
Mechanical — all except wheels, steering and brakes

(35.7 x (0.032 - 0.008 - 0.004)) = (34.1%)

Total 57.9%

Proportion of accidents that can be pushed

Assumed to be half, i.e., 50%

Proportion of all one-lane incidents that can be pushed

$$\frac{0.17}{0.94} \times 0.5 + \frac{0.77}{0.94} \times 0.579 = 56\%$$

16 The normal time it takes to perform the activity indicated in entry 14, e.g., police response time or other response time. The police response time is zero for pushable accidents. The other response time is the average time a police vehicle will need to reach the incident. This can range from zero to 10 minutes when a police vehicle is 10 miles away. An average value of 5 minutes is indicated.

17 List the product of entry 15 and entry 16.

18 This represents the percentage of incidents not affected by the conditions indicated in entry 14 and computed as 100 in entry 15.

19 The nominal time required for equipment to respond to a non-pushable incident. These values have to be determined for each segment on the basis of the local conditions. For the example, values were shown in table 3.1.

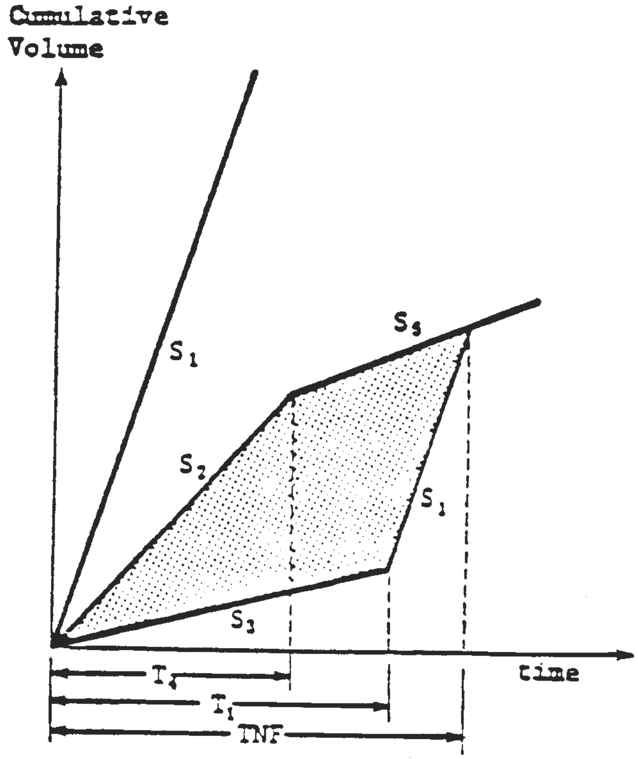
20 List the product of entry 18 and entry 19.

21 These columns are used in the same way as entries 15, 16, and 17 if an additional mode is being considered.

22 List the sum of the previous products yielding the weighted service time for each mode.

Section 3 of DCW (Base Case)

This section catalogues the traffic flow characteristics required for input to the delay calculations. The variables S_1, S_2, S_3, S_4, and T_4 are defined in figure 3.2. The situation envisaged in the example problem is illustrated in figure 3.7.



Notes:

For the 10-mile segment scenario, the algebraic variables have the following values:

S_1	=	5,500 vehicles per hour
S_2	=	5,000 vehicles per hour
S_5	=	2,500 vehicles per hour
S_3	=	2,700 vehicles per hour for an in-lane incident
S_3	=	4,600 vehicles per hour for a shoulder accident
T_4	=	60 minutes
TNF	=	Total time to normal flow — see text
T_1	=	Duration of the incident — see text

Figure 3.7. Delay Conditions for 10-Mile Segment

(23) From table 3.1 for three-lane facility.

(24) From table 3.2, the initial demand flow rate prior to the incident.

(25) From table 3.2, this represents the demand flow rate after the peak period if the incident extends beyond the peak period being evaluated.

(26) From table 3.1, this shows the bottleneck flow rates for an accident on the shoulder and an incident in a lane.

(27) T_4 is the duration of the initial demand flow rate S_2. However, for calculation purposes, the expected value of the duration (or one-half of it) is used because of the random nature of the incident occurrence. That is, there is an equal probability of the incident occurring during any minute of the period. An incident lasting 45 minutes and occurring during the first hour of the 2-hour rush period would be correctly evaluated at the S_2 flow rate. However, the same incident occurring during the last half hour of the 2-hour rush period would have to be evaluated under both S_2 and S_5 flow conditions. Using the expected value of the duration of the initial flow removes the need to have prior knowledge of when the incident occurs and facilitates evaluation. In the example, half of the 2-hour rush period, i.e., 60 minutes is used.

Section 4 of DCW (Base Case)

In this section of the form, the incident service times are computed for each detection method. T_1 is defined in figure 3.2.

(28) The number of shoulder accidents refers to those accidents that occur on the shoulder or that occur in-lane and make their way to the shoulder before a response vehicle arrives. This value is based on a calculation of the total number of incidents and a determination of those shoulder incidents that cause delay.

The total incident rate for this segment is based upon the incident formula. Total annual incidents = 200 incidents per million vehicle-miles (MVM) x 5,000 vehicles per hour x 10 miles x 4 hours per day x 250 days per year = 10,000 incidents per year. Note that 4 hours per day were used to capture both peak periods since the peak-period flow rates are nearly identical; doing this reduces the evaluation effort by 5 percent. Shoulder accidents for this entry are response-required accidents. From figure 2.5 (the general incident tree), 10,000 x .96 x .042 x .577 = 233 accidents.

(29) T_1 is the time that elapses before some change takes place in the incident situation that either enables some increase in flow to take place or temporarily stops the flow altogether. In the case of a shoulder accident, the impediment to flow exists until the incident has been cleared. Thus the value of T_1 is the sum of the detection time, the response time and the clearance time. For a shoulder accident, the response time is that time required to get the necessary emergency vehicles to the incident; that is, the nominal value. This is the same for all detection modes. Hence, only one value of T_1 need be calculated which is indicated by the deletion of police in the line entitled ALL/POLICE.

(30) Detection time from entry 13, response time from entry 19, clearance time is based on historical evidence for the particular area and was shown in table 3.2.

31 Derived in same way as entry 28, but this case is for incidents blocking one lane (i.e., one-lane accidents and one-lane disablements). Therefore, values are:

10,000 (0.04 x 0.711 x (0.213 x 0.799 + 0.787 x 0.979)) = 267 incidents

32 In this case, T_l has a different value for incidents detected by the police since pushable incidents have a zero response time. Thus the response times for police and others are taken from entry 22. The detection and clearance times remain as in entry 30.

Section 5 of the DCW (Base Case)

This section determines the delay caused by incidents using the equations given previously and noted below by equation number.

33 The delay per shoulder incident is computed directly as there is only one value for T_l in entry 29. Referring back to figure 3.3 and figure 3.7, the example problem will correspond either to Condition 1: Simple Blockage or Condition 4: Revised Demand, depending upon whether the end of the rush period is reached before the incident is cleared and flow fully restored. For the revised demand case to apply, the value of TNF must be greater than the value of T_4. For a simple blockage, the value of TNF is given by:

$$\mathrm{TNF} = \frac{T_l\,(S_l - S_3)}{S_l - S_2}$$

$$\mathrm{TNF} = \frac{57(5550 - 4600)/60}{(5550 - 5000)/60}$$

$$\mathrm{TNF} = 98.5 \text{ minutes}$$

Thus, TNF is greater than T_4, which is 60 minutes. The delay condition is, therefore, one of revised demand, and the delay per incident is given by equation **(9)** and the corresponding value of TNF by equation **(10)**. (Note that units must be consistent.) These equations yield values of 219 vehicle-hours of delay per incident and a time of 67 minutes until normal flow resumes.

34 The delay for in-lane incidents is made up of the components corresponding to the values of T_l in entry 31. For the in-lane case, the bottleneck flow is reduced to 2,700 so it is again clearly evident that the revised demand case applied. Thus, the delay and TNF are obtained by substituting into equations **(9)** and **(10)** giving values of 1,692 vehicle-hours of delay and 96 minutes when T_l = 51 minutes, and 1,801 vehicle-hours of delay and 100 minutes when T_l = 54 minutes.

35 The number of incidents/year is given by entry 28 for shoulder accidents and by entries 7, 9, and 31 for one lane incidents.

36 37 The delay/year is given by the product of the number of incidents and the delay per incident for each category. The base delay is the sum of all categories in annual vehicle hours (AVH).

38 40 These are used when an option to the base situation is being evaluated (see later).

Review of Base Situation

The completion of the DCW for the base case should be followed by a review of all the assumptions and the reasonableness of the evaluation results in the light of experience in the area under examination. Having completed this, the base case results are examined to determine whether any improvements can be made bearing in mind the options available, which were discussed earlier. Not all options are available or applicable to all areas, and the best option for one situation will not always be the best for another area even though the theoretical problem may be the same.

Thus, the engineer has to combine a logical examination of the evaluation results with his own knowledge of what is feasible in his own area. Taking the example, it is clear from examining the DCW and figure 3.1 that the greatest amount of delay is attributed to detection (27 minutes), followed by clearance (20 minutes). Response time is 50 percent less than these values. In this particular example, it appears that the prudent analyst would first examine options that would improve detection time, since it appears to be the greatest source of delay. (Other evaluations may have response time or clearance time as the greatest source of delay.) Later, when additional information is brought into this decision process (particularly information on option cost), refinements can be made.

INCIDENT DETECTION[1]

This part of the chapter details the mathematics that enable the engineer to estimate expected detection times. Techniques for detecting incidents are described in chapter 10 of this volume. Chapter 6 of volume 3 deals with on-site incident management.

Estimating Detection Time

Detection time is measured from the moment an incident occurs until the time that it has been seen by, or reported to, an official agency with incident management responsibilities. For planning purposes, detection is assumed to have taken place when one of the following events has occurred:

- The incident is observed by a police patrol, service patrol, or other part of the official incident management system
- The incident is observed by a passing motorist, and information is relayed to the official incident management agency (often to a police dispatcher) by telephone, callbox, CB radio, or through other means
- A motorist involved in the incident makes his way to a telephone or callbox and informs the official incident management agency
- An electronic incident detection device completes its verification steps and alerts the system monitor by means of an alarm or other warning device

Thus, detection time includes related activities such as recognition and notification. Subsequent actions undertaken by the incident management agency are considered to contribute to response time rather than detection time.

For planning purposes, a simple incident detection model has been developed to estimate expected detection time.

[1] The reader performing an initial evaluation of the existing system may wish to skip the next two sections (Incident Detection and Analysis of Options) and return to them later, when performing an alternatives analysis.

The model assumes that the various detection methods possible on a given freeway can all be represented as independent patrols, regardless of whether they actually are patrols. Another assumption of this model is that the headways between consecutive patrols are randomly distributed over a certain range. This appears to be a counterintuitive assumption, as an initial consideration of a police patrol operating on a regular beat suggests that a constant headway is more appropriate. Limited research(2) and interviews conducted with law enforcement officials confirm, however, that police officers on freeway patrol typically encounter a variety of random events (including incidents) such that the actual headway between patrols does not fit a regular pattern.

Detection time is measured from the first detection by any method that is described here as a mode. Because different types of responses may be required depending on the detection method, the probability of first detection by each detection mode must be considered in estimating detection time. The expected detection time for all modes acting together is used in computing the detection component of the total incident duration T_1.

The simple detection model expresses the probability of detection by a given mode, P_1, in terms of that mode's effective patrol frequency, R (which is defined as 1/average headway, or $1/H_1$), and a reporting delay, A, which will occur if the incident is detected by a mode other than a police patrol. Detection by a police patrol is assumed to have occurred as soon as an officer observes the incident, whereas other modes must relay reports to the police in order to complete the detection process.

Figure 3.8 is a graphic representation of the described detection model for three modes. In this case, R_1 represents the effective frequency of the first mode, a normal police patrol for which there is no reporting delay; hence, $A_1 = 0$.

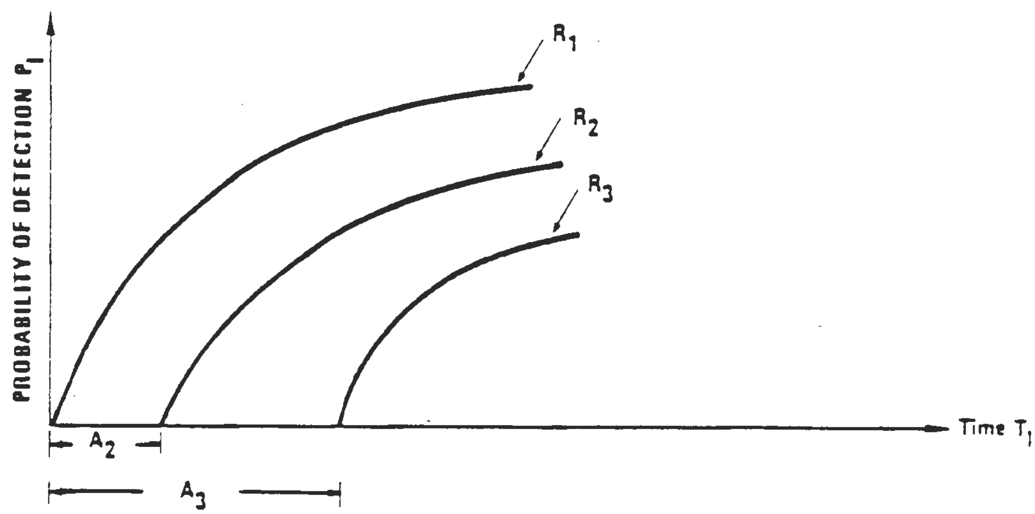


Figure 3.8. Detection Model for Three Modes

The mathematical expressions for the "three-mode model" for incident detection can be derived in a straightforward manner from probability theory and the properties of exponential distributions. For the general three-mode case with delays in reporting for all modes, the following expressions may be derived for the probability of first detection by each mode:

$$P_1 = 1 - \left[\frac{R_2}{R_1 + R_2}\right] e^{-c} \tag{11}$$

$$- \left[\frac{R_1 R_3}{(R_1 + R_2)(R_1 + R_2 + R_3)}\right] e^{-b}$$

$$P_2 = \left[\frac{R_2}{R_1 + R_2}\right] (e^{-c} - e^{-b}) \tag{12}$$

$$+ \left[\frac{R_3}{R_1 + R_2 + R_3}\right] e^{-b}$$

$$P_3 = \left[\frac{R_3}{R_1 + R_2 + R_3}\right] e^{-b} \tag{13}$$

$$c = R_1(A_2 - A_1) \tag{14}$$

$$b = R_1(A_3 - A_1) + R_2(A_3 - A_2) \tag{15}$$

Where:

P_1, P_2, P_3 are the probability of first detection by modes 1, 2, 3

R_1, R_2, R_3 are the frequency parameters of the negative exponential distribution for the three modes

A_1, A_2, A_3 are the reporting delays for the three modes A_3, A_2, A_1

The use of the three-mode model requires the estimation of certain parameters. Normally, the frequency parameter for the first mode, assumed to be a police patrol, would be known; values of A_1 would be assumed, with $A_1 = 0$ for a police patrol; and values for P_1 would be estimated from dispatchers' logs or experience. If these parameters are known, the remaining analysis will require estimation of the other two frequency parameters, R_2 and R_3, and the computation of the expected detection time, EX. The following expressions can be derived from the above equations and relations for expected values of joint exponential distributions:

$$R_2 = \frac{R_1 P_2}{P_1 - 1 + e^{-c}} \tag{16}$$

$$R_3 = \frac{(R_1 + R_2)P_3}{e^{-b} - [P_3]} \tag{17}$$

$$EX = A_1 + P_1/R_1 \tag{18}$$

Where:

EX is the average reporting delay for all modes

If the first mode is a police patrol with no assumed reporting delay, the expected value of detection time becomes simply P_1/R_1 or $P_1\ H_1$, where H_1 is the mean headway between patrol vehicles.

For an evaluation, the expected value for the current situation would be computed directly from equation (18) above. The values of the exponential coefficients would then be computed from equations (14) and (15), with the former being substituted into equation (16) to obtain the frequency parameter for the second mode, or pseudo-patrol. The results of (15) and (16) are then used in equation (17) to obtain the values for the third mode.

Several different policy assumptions can then be made. Normally, the values of the frequency parameters for the second and third modes are taken as constant. If a revised police patrol were evaluated, the value of R_1 would be changed accordingly. Revised values of the exponential coefficients would then be computed from equations (14) and (15), and a revised probability of detection by the police patrol from equation (10). The new expected value can then be computed directly from equation (18).

This three-mode model is quite useful for almost any realistic incident management system. However, the resulting calculations require the use of exponentiation and are rather cumbersome. Unfortunately, the lack of separability in the relationships makes the preparation of simple nomographs impossible.

In many situations, a two-mode detection model is sufficing. This has simple properties, especially if the first mode is assumed to be a normal police patrol. The resulting equations are:

$$P_1 = 1 - \left[\frac{R_2}{R_1 + R_2}\right] e^{-c} \tag{19}$$

$$P_2 = 1 - P_1 \tag{20}$$

$$c = R_1 A_2 \tag{21}$$

$$R_2 = \frac{R_1 P_2}{P_1 - 1 + e^{-c}} \tag{22}$$

$$EX = P_1 / R_1 \tag{23}$$

The application of these models is described later. However, three basic parameters must be known or estimated in order to make use of the detection model. The first is the average headway or patrol frequency of the "base" patrol, usually standard police patrols. If more than one organization operates a patrol on a particular freeway section, the frequency of all units must be determined.

The second parameter is the probability that an incident will be detected first by a base patrol unit. This can be estimated from an analysis of dispatching logs or from discussions with experienced dispatchers. If more than one base patrol is in operation, it is assumed that the detection probabilities assigned to each will be proportional to their relative patrol frequencies.

Finally, the reporting delay for modes other than the base police patrol must be estimated. No rigorous method exists for determining this value but, fortunately, the detection model is not overly sensitive to this parameter. A typical set of estimated values and the reporting conditions they might simulate are:

- <u>2 minutes</u> — well coordinated CB monitoring by public agencies or private citizen group base stations; closely spaced, readily usable callboxes
- <u>5 minutes</u> — loosely coordinated CB monitoring; typical callbox system
- <u>10 minutes</u> — little CB reporting of incidents; relatively dense urban area

with closely spaced interchanges and commercial activity with telephone nearby

- 15 minutes – suburban area with longer interchange spacing and few readily available telephones
- 20 minutes – fairly remote areas with limited access to telephone

These reporting delays are intended to represent averages for reporting done by passing motorists and involved motorists.

In general terms, the incident detection estimation procedure is used first to estimate the expected detection time under the existing or base case situation. Then, incident management options that significantly reduce detection time are evaluated by changing the base and/or other patrol frequency and computing a revised detection time.

ANALYSIS OF OPTIONS USING DCW

Option choice must occasionally involve consideration of what agency will take the lead role. For example, the benefit to a highway agency engineer of evaluating an option such as increasing the police patrol frequency is unclear. This is an option that the agency has little or no control over (unless it is taking an advocacy position and is trying to convince the police agency to increase the patrol frequency). Similarly, police evaluation of the CCTV option for large segments may be questionable, due to cost considerations. In the illustrative discussion that follows, the evaluation is being undertaken by a team composed of all participants in the freeway environment, with day-to-day incident management activities being conducted by police personnel in close coordination with highway agency personnel.

After it has been decided which agency will take the lead role, the option list must be examined to determine the total number of options that can impact detection time. An examination of the options suggests that there are seven options discussed that relate to finding a solution to detection problems. These are as follows:

- Increased police patrol frequency
- CB radio monitoring
- Callboxes
- Peak-period motorcycle patrol
- Service patrol
- Stationary response vehicle
- Aircraft

For the sample problem, three options have been evaluated. These have been selected to demonstrate the range of computational techniques and procedures rather than to examine those with greatest merit.

Three options are evaluated below:

- Increased police patrol frequency
- CB radio monitoring
- Service patrol

The DCW for each option is described using a similar format to that adopted previously. However, only changes to the base case DCW are labeled and described. The descriptions only cover new issues arising from the evaluation of the option. Hence, if an entry is not completely understood, it may be necessary to review the corresponding entry on the DCW for the base case (figure 3.5).

Increased Police Patrol Frequency

Section 1 of the DCW – This option involves increasing the police patrol frequency (reducing headways) by placing additional police patrol vehicles on the road or by reducing the nonpatrol duties of the existing police officers. The assumption included in the evaluation is that the headway can be halved by increasing the number of patrol cars by one vehicle.

1. List option name (figure 3.9).

2. This entry reflects half the base headway of 90 minutes.

3. Half the headway as detection across the median is assumed.

4. The value of P_1 may be determined using the equations **(19-23)**, together with the information known about the base condition. The frequency parameter R_2 for the base and the option conditions will be the same. Thus, R_2 for the base can be computed from equation **(22)**. This will equal R_2 for the option and hence, equation **(19)** can be used to compute P_1 given also the option value for R_1, entry 3. Thus:

$$R_2 = \frac{(1/45)(0.4)}{0.6 - 1 + e^{-(1/45)(10)}}$$

$$= 0.02218$$

$$P_1 = 1 - \frac{0.02218}{(1/22.5) + (0.02218)} e^{-(1/22.5)(10)}$$

$$= 0.78 \text{ or } \underline{78\%}$$

5. $100\% - P_1 = 22\%$

6. Taken from equation **(23)** (or figure 3.6)

$EX = P_1/R_1 = 0.78/0.0444 = \underline{17.5 \text{ minutes}}$

Sections 2 and 3 of the DCW – These sections remain unchanged from the base condition.

Section 4 of the DCW – The value of T_1 is amended in entry 8 to reflect the revised detection time in entry 7 taken from entry 6.

Section 5 of the DCW – As the value of T_1 has changed, a check must be made to determine which delay condition now prevails. From equation **(4)**;

$$TNF = \frac{T_1(S_1-S_3)}{S_1-S_2} = \frac{48\ (5550-4600)/60}{(5550-5000)/60}$$

TNF = 83 minutes

As this is larger than T_4 (60 minutes), the revised demand condition applies and the delay per incident can be computed from equation **(9)** and TNF from equation **(10)** entry 9.

10. The number of incidents in each category is revised to reflect the new percentages in entries 4 and 5.

11. 12. The option delay is the sum of the delay per year for each category and is less than the base delay by 102,155 annual vehicle-hours entry 13.

14. 15. Both entries show an estimate of the annual cost of this option.

CB Radio Monitoring

Section 1 of the DCW – This option involves the installation of CB radios in the police vehicles operating in the base condition.

1. The percentage of incidents first detected by CB radio must be assumed. The 20 percent value on figure 3.10 was obtained from Mis-

DELAY COMPUTATION WORKSHEET

1

Segment 92-8 Time Period Am (EB), PM (WB) Lanes 3 Shoulders Y/N

System/Option Double Base Case Police Patrol (1)

P_1 (Police) = 78 (4) %, P_2 (Other) = 22 (5) %, A_2 = 10 Min.

Headway = 45 (2); R_1 = 1/22.5 (3) Expected Detection Time = 17.5 (6) Min.

2 WEIGHTED SERVICE TIME (Minutes)

TIME	% of Incid.	Nom. Time.	Product	% of Incid.	Nom. Time	Product	% of Incid.	Nom. Time	Product	TOTAL
Resp. - Police	56	0	0	44	10	4.4				4.4
Resp. - Other	56	5	2.8	44	10	4.4				7.2

3 DELAY COMPUTATION VARIABLES

S_1 (Veh/Hr)	S_2 (Veh/Hr)	S_5 (Veh/Hr)	S_3 (Veh/Hr) Shldr	S_3 (Veh/Hr) Lane	T_4 (Min.)
5550	5000	2500	4600	2700	60

4 INCIDENT SERVICE TIMES FOR EACH DETECTION METHOD (Minutes)

SHOULDER ACCIDENTS – No. 233

	Detection (7)	+ Response	+ Clearance	= TOTAL (3)
T_1: ~~ALL~~/POLICE	17.5	10	20	48
T_1: OTHER				
T_1: SPECIAL ______				

IN-LANE INCIDENTS – No. 267

	Detection	+ Response	+ Clearance	= TOTAL
T_1: ALL/POLICE	17.5	4	20	42 (3)
T_1: OTHER	17.5	7	20	44 (3)
T_1: SPECIAL ______				

5 TOTAL DELAY FOR EACH DETECTION METHOD (Vehicle-Hours)

	Shoulder Accidents: ~~All/Police~~	Shoulder Accidents: Other	Shoulder Accidents: Special	In-Lane Incidents: ~~All~~/Police	In-Lane Incidents: Other	In-Lane Incidents: Special
DELAY/INCIDENT	188	(9)		1364	1437	
NO. OF INCIDENTS/YEAR	233			(10) 208	(10) 59	
DELAY/YEAR (11)	43,804		(12)	283,712	84,783	

BASE DELAY: 514,454 – OPTION DELAY: 412,299 = 102,155 (13) A.V.H.

COST OF THIS OPTION $17,411 (14)

NOTES: 52 miles/day patrol, 20 miles/day access ~ 18,000 miles/yr. (15)

Cost = $1811 veh + $13,800 labor + 1,800 M+O

Figure 3.9. DCW for Increased Police Patrol Frequency

DELAY COMPUTATION WORKSHEET

1

Segment 92-8 Time Period Am (EB) Pm (WB) Lanes 3 Shoulders Y/N

System/Option CB in Police Vehicles (CBers report 20%) 1

P_1 (Police/CB) 2 = 70 %, P_2 (Other) = 3 30 %, A_2 = 10 Min.

Headway = N.A. 4; R_1 = 1/31 5 Expected Detection Time = 21 6 Min.

2 WEIGHTED SERVICE TIME (Minutes)

TIME	% of Incid.	Nom. Time	Product	% of Incid.	Nom. Time	Product	% of Incid.	Nom. Time	Product	TOTAL
Resp. - Police	56	0	0	44	10	4.4				4.4
7 Resp. - CB/Other	56	5	2.8	44	10	4.4				7.2

3 DELAY COMPUTATION VARIABLES

S_1 (Veh/Hr)	S_2 (Veh/Hr)	S_5 (Veh/Hr)	S_3 (Veh/Hr) Shldr	S_3 (Veh/Hr) Lane	T_4 (Min.)
5550	5000	2500	4600	2700	60

4 INCIDENT SERVICE TIMES FOR EACH DETECTION METHOD (Minutes)

SHOULDER ACCIDENTS – No. 233

	Detection	+ Response	+ Clearance	= TOTAL
T_1: ALL/~~POLICE~~	21 8	10	20	51 9
T_1: OTHER				
T_1: SPECIAL ______				

IN-LANE INCIDENTS – No. 267

	Detection	+ Response	+ Clearance	= TOTAL
T_1: ~~ALL/~~POLICE	21	4	20	45
T_1: OTHER	21	7	20	48 9
T_1: SPECIAL CB	21	7	20	48

5 TOTAL DELAY FOR EACH DETECTION METHOD (Vehicle Hours)

	Shoulder Accidents: ~~All/Police~~	Shoulder Accidents: Other	Shoulder Accidents: Special	In-Lane Incidents: ~~All/~~Police	In-Lane Incidents: Other	In-Lane Incidents: Special
DELAY/INCIDENT	200	10		1473	1583	1583
NO. OF INCIDENTS/YEAR	233			133 11	80	11 54
DELAY/YEAR 12	46,600		13	195,909	126,640	85482

BASE DELAY: 514,454 – OPTION DELAY: 454,631 = 59,823 14 A.V.H.

COST OF THIS OPTION $47 15

NOTES:

Figure 3.10. DCW for CB Radio Monitoring

souri State Highway patrol rural experience.

2 3 The percentage of incidents first detected by CB's was assumed at 20 percent. It must also be assumed from whence this 20 percent will be drawn. It is arbitrarily assumed that the 20 percent is derived by taking 10 percent from each existing mode. In other words, police patrol, other, and CB's are assumed to detect 50, 30, and 20 percent, respectively. Police and CB-detected percentages are combined because a CB-detected incident is assumed to be detected by a pseudo patrol. The two-mode model is discussed in this chapter and the pertinent variables can be calculated using equations (19-23).

4 This entry is not applicable because the percentage of incidents detected by police is already known.

5 The frequency R_1 can be derived from equation (19) using the value of P_1 from entry 2 and R_2 from the base condition as in entry 4 for the previous option. Since this is a complex process, it is worthwhile to construct a series of curves as shown in figure 3.11. Note, however, that these are only valid for one value of A_2. Enter figure 3.11 on the base case horizontal line and follow it to its intersection with the 70 percent isoquant. A vertical projection of the intersection to the time axis produces the indicated 31 minutes.

6 Taken from equation (23).

Section 2 of the DCW –

7 The weighted service times are the same as in the base case, but the CB-detected incidents are included in "other" in this instance.

Section 4 of the DCW –

8 T_1 is amended to reflect the revised detection time from entry 6.

9 CB incidents are shown separately but as there is still a need for a police response time equivalent to "other," the value of T_1 is the same as for CB and "other."

Section 5 of the DCW –

10 Delay per incident is calculated as for previous option.

11 The number of incidents reflects the percentages shown in entry 2, namely 50, 30, and 20.

12 List delay/year calculated as before.

13 14 The option delay is 454,631, a reduction of 59,823 annual vehicle-hours over the base situation.

Service Patrol

Section 1 of the DCW – This option involves a medium-duty wrecker patrolling the freeway at the same frequency and in addition to the existing base police patrol. A police officer or police cadet rides in the wrecker to handle police matters such as accident investigations. The wrecker is equipped with a police radio. (See figure 3.12.)

Entries 2, 3, 4, 5, and 6 are identical to previous CB radio monitoring option.

Section 2 of the DCW – The weighted service time must reflect the three types of incident detection situations.

8 From the incident tree in figure 2.5 it is evident that a wrecker can handle all incidents except medical emergencies and fire. Thus, subtracting from 100 percent the inci-

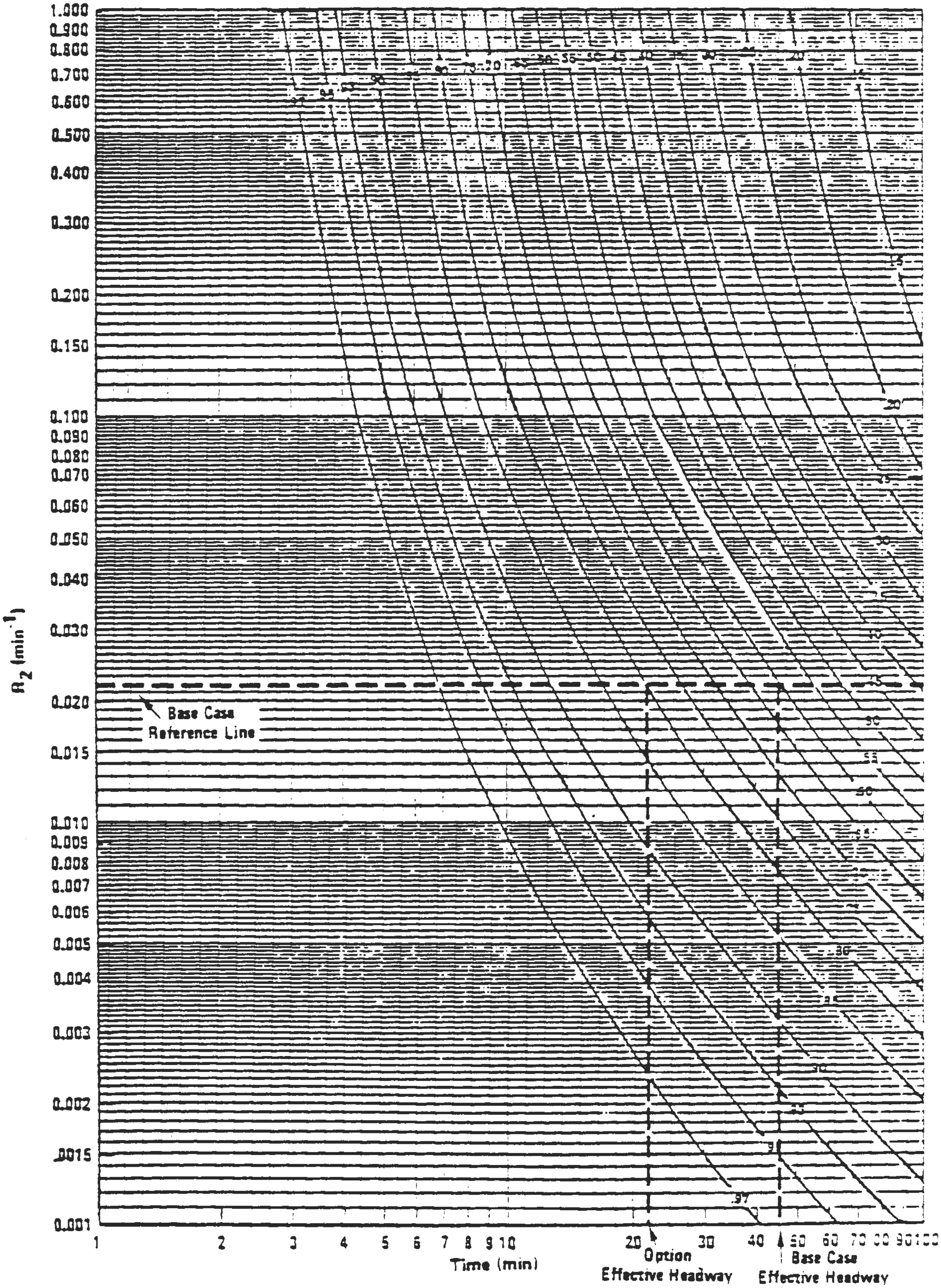


Figure 3.11. Detection Probability Estimation

DELAY COMPUTATION WORKSHEET

1

Segment 92-8 Time Period Am (EB), PM (WB) Lanes 3 Shoulders Y/N

System/Option mobile wrecker with Police officer to Double base case 1

P_1 (Police/wrecker) = 4 78 %, P_2 (other) = 5 22 %, A_2 = 10 Min.

Headway = 45 2; R_1 = 1/22.5 3 Expected Detection Time = 17.5 6 Min.

2 WEIGHTED SERVICE TIME (Minutes)

TIME 7	15 % of Incid.	16 Nom. Time	10 Product	8 % of Incid.	9 Nom. Time	10 Product	11 % of Incid.	12 Nom. Time	10 Product	TOTAL 13
Resp. - police	56	0	0	38	7	2.7	6	10	0.6	3.3
14 Resp. - wrecker	94	0	0	6	10	0.6		-		0.6
19 Resp. - other	56	5	2.8	38	7	2.7	6	10	0.6	6.1

3 DELAY COMPUTATION VARIABLES

S_1 (Veh/Hr)	S_2 (Veh/Hr)	S_5 (Veh/Hr)	S_3 (Veh/Hr) Shldr	S_3 (Veh/Hr) Lane	T_4 (Min.)
5550	5000	2500	4600	2700	60

4 INCIDENT SERVICE TIMES FOR EACH DETECTION METHOD (Minutes)

SHOULDER ACCIDENTS – No. 233

	20 Detection	+ 21 Response	+ Clearance	= TOTAL 22
T_1: ~~ALL~~ POLICE	17.5	7	20	45
T_1: OTHER	17.5	7	20	45
T_1: SPECIAL wrecker	17.5	0	20	38

IN-LANE INCIDENTS – No. 267

T_1: ~~ALL~~ POLICE	23 17.5	24 3.3	20	41 25
T_1: OTHER	17.5	6.1	20	44
T_1: SPECIAL Wrecker	17.5	0.6	20	38

5 TOTAL DELAY FOR EACH DETECTION METHOD (Vehicle Hours)

	SHOULDER ACCIDENTS ~~All~~/Police	Other	Special	IN-LANE INCIDENTS ~~All~~/Police	Other	Special
DELAY/INCIDENT 26	174	174	136	1327	1437	1216
27 NO. OF INCIDENTS/YEAR	91	51	91	104	59	104
28 DELAY/YEAR	15834	8874	12376	138008	84783	126464

BASE DELAY: 514,454 – OPTION DELAY: 386,339 29 = 128,115 A.V.H. 30

31 COST OF THIS OPTION $20,656 police Driver; $34,458 police pass; $27,800 rental

NOTES: 18,000 mi @ 24¢ = $2536 Veh + $13,800 labor (= $13,800 (driver))

4 hr x 250 days x $15/hr = $13,800 police passenger

Figure 3.12. DCW for Service Patrol Option

dents the wrecker cannot handle yields the required percentages. Use a procedure similar to the base case entry 15.

(9) The response time for the wrecker has been increased by 2 minutes over the 5 minutes for the police cruiser to account for its reduced mobility.

(10) Computed as previously.

(11) This represents the 6 percent of incidents requiring a fire or ambulance response.

(12) List response time from table 3.2.

(13) Compute as previously.

(14) This category is necessary to distinguish between wrecker-detected incidents that can be dealt with immediately and the 6 percent requiring fire or ambulance assistance.

(15) See entry 8.

(16) Wrecker can handle these incidents immediately—equivalent to pushable incidents for police cruisers.

(17) See entry 8.

(18) From table 3.2, worst case of fire and ambulance response.

(19) Fifty-six percent of "other"-detected incidents dealt with by police cruiser, 38 percent by wrecker, 6 percent fire and ambulance.

Section 4 DCW — Incident service times adjusted to reflect third mode and amended response times.

(20) Taken from entry 6.

(21) It is assumed that wrecker can handle all shoulder incidents. See entry 9 for response times.

(22) Use new totals.

(23) Taken from entry 6.

(24) Taken from entry 13.

(25) Use new totals.

Section 5 of the DCW — Delay computed as previously. Check TNF value for lowest value of T_1 which is 38 minutes.

$$TNF = \frac{38(5550-4600)/60}{(5550-5000)/60}$$

$$TNF = 66 \text{ minutes}$$

Thus, all cases are revised demand condition.

(26) Taken from Equation (9).

(27) Incidents allocated by detection mode as entries 4 and 5; police-detected incidents equally divided between cruiser and wrecker.

(28) Previous calculation, (entry 26 x entry 27).

(29) (30) Total option delay 386,339 AVH which reflects a reduction of 128,115 AVH over the base condition.

(31) An estimate of the option cost.

OTHER PERFORMANCE MEASURES

Several measures other than vehicle-hours of delay can be used to analyze the congestion and queueing situations that can occur following an incident. They include the time for normal traffic flow to resume (TNF), the total number of vehicles delayed by the incident (VT), the average delay per vehicle (AD), the maximum length (expressed in number of vehicles) of the queue (QM), and the length of the maximum queue (QL) expressed in miles. Unlike total delay, however, these measures are not readily aggregated to determine an overall measure of effectiveness. They are primarily useful in evaluating specific incidents or types of incidents under certain conditions.

Time for Normal Flow to Resume (TNF)

Since it directly reflects the length of time that congestion will be present on the freeway following an incident, TNF has a direct interpretation as an evaluation measure. Under certain conditions TNF can also be used to assess the adequacy of a given incident management strategy. For example, in evaluating incidents occurring prior to peak periods, it is important that the TNF be short enough to avoid running over into the peak period and causing significant increases in congestion.

Total Vehicles Delayed (VT)

VT is closely related to TNF and is defined as the number of vehicles passing the incident from the moment the incident occurs until normal flow resumes. Like delay and TNF, VT is dependent on traffic flow rates and incident duration.

As illustrated in figure 3.3, VT can be estimated for different delay conditions. VT is determined by the demand flow rates, S_2 and S_5, and the TNF. If a demand reduction does not occur, or if the incident is cleared before a demand reduction occurs, VT can be computed simply as the product of the basic demand flow rate, S_2, and the TNF, or:

$$VT = S_2\,(TNF) \quad \text{for } T_4 = 0 \text{ or } TNF \quad T_4 \qquad (24)$$

If a demand reduction occurs before normal flow is restored, the number of vehicles delayed can be calculated from the demand flow rate, S_2 (until the time, T_4, of the demand reduction) and the reduced flow rate, S_5, for the remaining time interval ($TNF - T_4$), or:

$$VT = S_2\,T_4 + S_5\,(TNF - T_4)$$

$$\text{for } 0 \quad T_4 \quad TNF \qquad (25)$$

Average Delay Per Vehicle (AD)

AD is another simple direct measure of the impact of incidents on freeway traffic. It can be derived from the estimated total delay and the number of vehicles involved (VT).

As mentioned before, however, delay has the basic advantage of being additive across different incident types, time periods, or freeway sections. AD cannot be aggregated in this manner. An approximate measure of overall average delay can be computed, however, by aggregating the total delay and the total number of vehicles involved for a given set of incident types.

Maximum Extent Of Queue (QM)

QM is an excellent measure of incident impact because it reflects the extent of congestion under the worst conditions. QM is difficult to measure precisely due to the stopping, starting, lane changing, and other movements of vehicles in a congested queue. These factors cause the actual number of vehicles in a queue at any given time to be highly variable. A simple estimate can be made as the difference between the number of vehicles that would have passed the incident site under free flow and the actual number that do pass the site under reduced flow conditions.

The maximum queue will occur at one of the times when flow rates change, as shown for the general case in figure 3.13. In this figure, four potential times could represent maximum queue build-up, although in reality no more than three are relevant for any given combination of times.

The critical element in computing the maximum queue is the time of the demand reduction, T_4, relative to the other action times, T_1, T_2, and T_3. A total of five cases can occur, with T_4 either not occurring or falling between the other times.

As can be seen from figure 3.13, maximum queue can be computed from the normal flow minus the restricted flow at a given point in time. This can be computed, in general terms, from the five flow rates multiplied by appropriate time intervals, depending upon the specific conditions. The general form of the equation is:

$$QM = T_a S_2 + T_b S_5 - T_c S_3 - T_d S_4 - T_e S_1 \qquad (26)$$

The intervals can be estimated by inspection of a sketch of the particular situation, such as shown in figure 3.13.

Alternatively, the appropriate values of T_a, T_b, T_c, T_d, and T_e for different cases depending on the value of T_4 are summarized in figure 3.14. For the cases illustrated previously in figure 3.3 the equation for each condition is as follows:

Condition 1: Simple Blockage

$$QM = T_1 S_2 - T_1 S_3 \qquad (27)$$

Condition 2: Short Term Closure

$$QM = (T_1 + T_2) S_2 - T_1 S_3 \qquad (28)$$

Condition 3: Adjusted Bottleneck

$$QM = (T_1 + T_3) S_2 - T_1 S_3 - T_3 S_4 \qquad (29)$$

Condition 4: Revised Demand

$(T_1 > T_4)$ and $(S_5 > S_3)$

$$QM = T_4 S_2 + (T_1 - T_4) S_5 - T_1 S_3 \qquad (30)$$

As noted previously for total delay, tabulated values of QM may be found in volume 6 of reference 1.

Maximum Queue Length (QL)

The impact of QL may be expressed as the maximum distance that the queue will extend upstream of an incident. This measure is useful in determining the impact of the incident on other sections of the freeway and the potential impact of an incident on highways that intersect with the freeway at upstream interchanges.

Maximum queue length is computed as follows:

$$QL = \frac{QM}{\text{No. of Lanes} \times \text{VLM}} \qquad (31)$$

Fuel Consumption

When an incident causing serious congestion occurs on a freeway, considerable quantities of fuel are wasted as vehicles idle, start, and stop. Research is currently being conducted to measure the amount of fuel consumed under a variety of highway conditions. Among the factors involved in this determination are: the mix of vehicles, their acceleration and deceleration characteristics within the

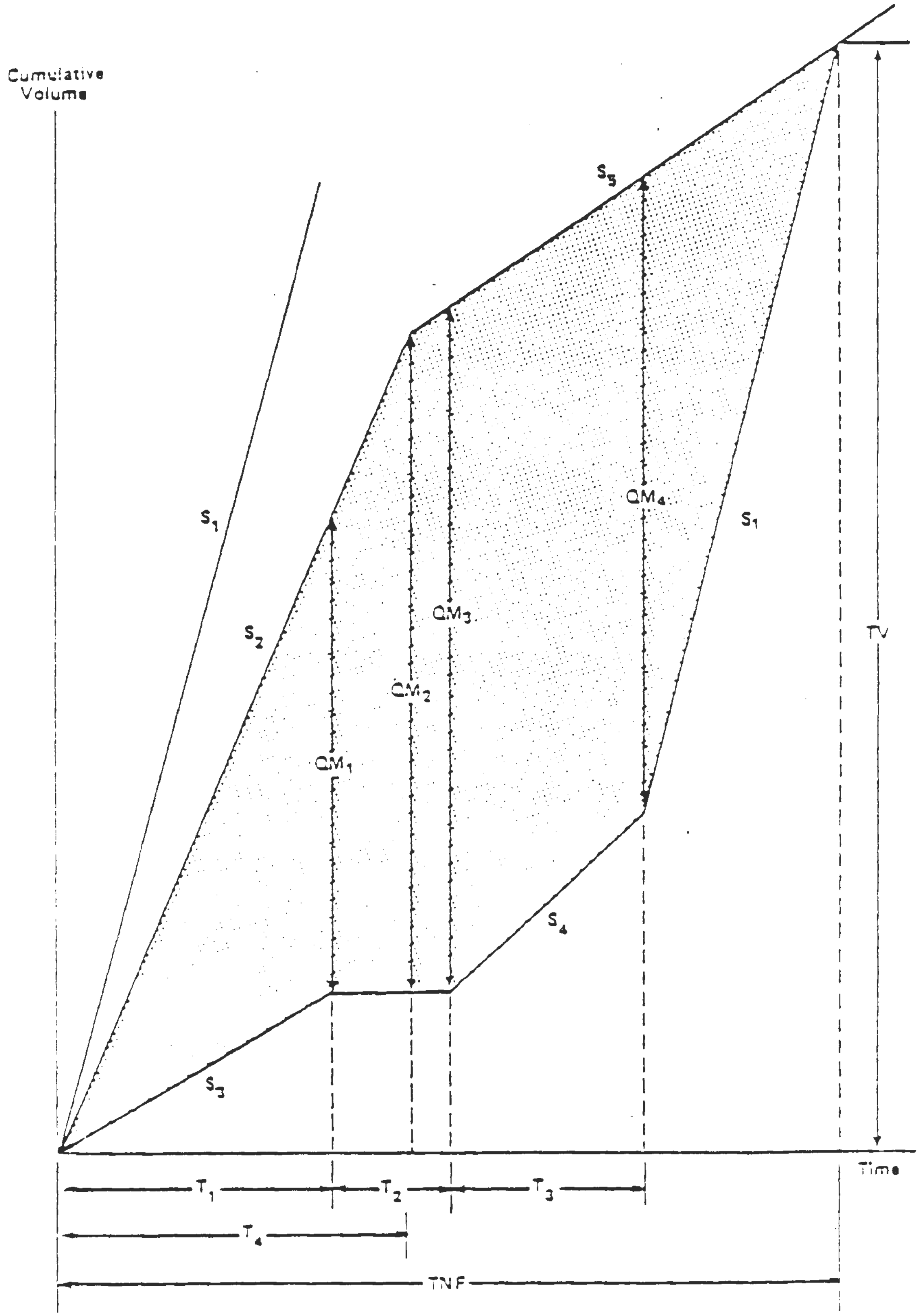


Figure 3.13. Maximum Queue Conditions

$$QM = T_a S_2 + T_b S_5 - T_c S_3 - T_d S_4 - T_e S_1$$

CASE	RANGE OF T_4	CRITICAL THF	T_a	T_b	T_c	T_d	T_e
I	0	$T_1 \leq THF$	T_1	0	T_1	0	0
II	$0 < T_4 \leq T_1$	$(T_1 + T_2) \leq THF$	T_4	$T_1 + T_2 - T_4$	T_1	0	0
		$(T_1 + T_2 + T_3) \leq THF$	T_4	$T_1 + T_2 + T_3 - T_4$	T_1	T_3	0
		$T_4 \leq THF$	T_4	0	T_4	0	0
III	$T_1 < T_4 \leq (T_1 + T_2)$	$(T_1 + T_2) \leq THF$	T_4	$T_1 + T_2 - T_4$	T_1	0	0
		$(T_1 + T_2 + T_3) \leq THF$	T_4	$T_1 + T_2 + T_3 - T_4$	T_1	T_3	0
IV	$(T_1 + T_2) < T_4 \leq (T_1 + T_2 + T_3)$	$(T_1 + T_2) \leq THF$	$T_1 + T_2$	0	T_1	0	0
		$(T_1 + T_2 + T_3) \leq THF$	T_4	$T_1 + T_2 + T_3 - T_4$	T_1	T_3	0
		$T_4 \leq THF$	T_4	0	T_1	$T_4 - T_1 - T_2$	0
V	$T_4 < (T_1 + T_2 + T_3)$	$(T_1 + T_2) \leq THF$	$T_1 + T_2$	0	T_1	0	0
		$(T_1 + T_2 + T_3) \leq THF$	$T_1 + T_2 + T_3$	0	T_1	T_3	0
		$T_4 \leq THF$	T_4	0	T_1	T_3	$T_4 - T_1 - T_2 - T_3$

Figure 3.14. Maximum Queue Parameters

queue, and the extent of the congestion over time and distance.

The derivation and application of a relationship accounting for these and other factors contributing to fuel consumption is both extremely difficult and inappropriate in the context of a planning analysis. Therefore, a simpler approach is recommended which assumes that the fuel wasted by delayed vehicles is reasonably estimated by the fuel that is consumed during the incident's duration. This approach further assumes that fuel is consumed at a constant rate for each vehicle-hour of delay. A conservative estimate of this rate is the average idle fuel flow rate of a typical mix of vehicles. One study (3) has shown that the idle fuel flow rate of a late-model vehicle of intermediate size is approximately .666 gallons per hour. This rate can be used to obtain a simple estimate of the fuel consumption impact of a given amount of estimated incident-related delay.

Fuel Consumed = 0.0111 Gal/Veh-min of delay (32)

Illustrative Example: Additional Evaluation Measures

Additional evaluation measures include a number of items that are derived from the basic delay calculation but which are often more easily understood by a non-technical audience. The computation of these is discussed in this section for the sample problem considered earlier.

As for delay, the initial calculations must be made for the base case and these have also been arranged on a pro forma to assist the evaluation process. This is shown in figure 3.15 for the base case and the numbered entries are discussed below.

Additional Evaluation Measures – Base Case

1 2 4 Taken from appropriate DCW, figure 3.11 for base case.

3 Each incident type (i.e., each value of T_1) must be evaluated on a separate form.

5 Taken from entry 36, figure 3.5.

6 Taken from entry 34, figure 3.5.

7 Taken from equation (10) using values from base case DCW (see also discussion of entry 34 on figure 3.1).

8 The total number of vehicles delayed from equation (24) or (25) for example equation (25) is appropriate;

$$VT = S_2T_4 + S_5\,(TNF - T_4)$$
$$= (5000)(60/60) + 2500\,(96-60)/60$$
$$= \underline{6500} \text{ vehicles}$$

9 Average delay vehicles; entry 6/8.

10 The appropriate equation is (27) as T_1 is less than T_4. Thus:

$$QM = T_1S_2 - T_1S_3$$
$$= 51/60 \times 5000 - 51/60 \times 2700$$
$$= \underline{1955}$$

11 Queue length from equation (31);

$$QL = 1955/3 \times 150 = \underline{4.34} \text{ miles}$$

12 Fuel consumption from equation (32)

$$\text{Fuel} = 270720 \times 0.666$$
$$= \underline{180.480} \text{ gallons}$$

13 Air pollution from chapter 4 of this volume

14 19 Illustrative computation of cost of delay

Additional Evaluation Measures – Service Patrol Option

Figure 3.16 shows the computation of the additional evaluation measures for the service patrol option. The procedure is the same as for the base case but the input values are taken from figure 3.12.

To complete the evaluation of other measures, figure 3.16 should be completed for the other incident types (that is, values of T_1). Only the measures of total delay, fuel, pollution, and delay costs can be meaningfully aggregated access incident categories.

EVALUATING NON-QUANTIFIABLE OPTIONS

Ideally, it would be preferable to evaluate all options by means of the DCW. However, some options require a special application of the DCW, and other options cannot be evaluated with the DCW at all. Such options have been defined as non-quantifiable options. This does not mean that the merits of these options cannot be determined, but rather that these options require special attention.

Examples of these options include the use of the dispatchers' manual, hazardous materials manual, policy/highway department relationship, and communications training. Illustratively, the merits of a good dispatchers' manual are that the police dispatchers' search time for a proper incident response in minimized. However, most dispatcher manuals contain frequently called numbers. Therefore, the benefits of the dispatchers' manual option are prevalent when unique incidents occur that require specialized response. To date, no information has been published regarding the number or frequency of such incidents, although it is known that such incidents occur. Thus, it is impossible to determine the benefits of the option using the DCW procedures. Similar evaluative problems could be cited for the other options noted above.

An alternate set of procedures has been developed to evaluate such options. This procedure requires more judgment than working through the DCW and may produce results having somewhat less precision.

The alternate procedures begin by identifying the fact that the option is non-quantifiable. Once this fact has been determined, one of three methods can be used to evaluate the option.

Method 1

The first method is subjective and can best be described by an example. For this illustration, it is assumed that the police are the lead agency and that consideration is being given to implementing the service patrol option. To implement the service patrol option, the police agency must establish a working relationship with the Department of Transportation (DOT). Thus, the highway agency relationship, would probably be implemented first. Although the highway agency relationship option is non-quantifiable, it would contribute to the success of implementing the service patrol. Therefore, an evaluation of this option as prescribed by Method 1 would result in the following subjective statement: "The merits of the highway agency relationship option are that it would contribute to the success of the service patrol option which, if implemented, is estimated to produce a savings of X vehicle-hours of delay."

Method 2

Method 2 involves creating a performance measure that is unique to the option being evaluated. For example, the hazardous manual option is a non-quantifiable option that results in the establishment of hazardous materials manuals that describe handling procedures. One performance measure might be the avoidance of loss of life through the use of correct procedures provided by the manual. Another performance measure might be the avoidance or limitation of spills through proper procedures outlined in the manual. Thus, the second method of evaluation would involve the engineer using judgment and professional skill to create performance measures tailored to a specific option and to local conditions.

Method 3

The third method for evaluating these options would involve determining if the option being evaluated could potentially affect the bottleneck flow rate, clearance time, or demand flow rate. For example, suppose the traffic operations training option was under consideration. The purpose of this option would be to increase the on-site incident removal expertise of the police officer. This option has the potential of reducing clearance time and of increasing bottleneck flow (flow past the incident). For illustrative purposes, suppose that it is hypothesized that the bottleneck flow rate could be improved by X percent by implementing traffic operations training. This improvement quantification lends itself to special DCW calculations as indicated in table 3.3.

Table 3.3 has been created to evaluate options whose benefits are estimated to have an improvement in clearance time, bottleneck flow rate, or demand flow rate. From above, the X percent improvement (let us say 5 to 10 percent) would be determined from table 3.3 to produce an 8 to 16 percent improvement in delay reduction during the rush hour. Table 3.3 has been developed from the 10-mile segment base case discussed previously.

If local conditions are such that table 3.3 is not applicable, a similar table could be developed, using the procedures described earlier.

In addition to being able to estimate the merits of certain options, this table also provides an indication of what types of options would provide the greatest benefits for local application. For example, table 3.3 indicates that a hypothetical 25 percent improvement in bottleneck flow rate would produce an estimated 38 percent reduction of delay during the peak period or an estimated 38 percent reduction during off-peak periods. Obviously, other local base conditions may produce different results with these types of improvements.

COST EFFECTIVENESS

The cost of implementing each option should be computed using local data and current values. Some typical hardware cost values that can be used in a preliminary analysis are contained in chapter 15 of volume 3.

When the costs are available, a figure similar to that shown in figure 3.17 can be constructed to demonstrate the relative cost effectiveness of different options. This shows that for the example considered, CB radios in police vehicles achieves 47 percent of the delay savings at 0.2 percent of the cost of the mobile wrecker option.

Table 3.3. Reductions in Delay Due to Improved Service Parameters

SERVICE IMPROVEMENT	DELAY REDUCTION	
	PEAK FLOW CONDITIONS (%)	OFF-PEAK CONDITIONS (%)
5% Improvement in Bottleneck Flow Rate (S_3)	8	21
10% Improvement in Bottleneck Flow Rate (S_3)	16	40
25% Improvement in Bottleneck Flow Rate (S_3)	38	88
25% Reduction in Clearance Time (8% Overall Reduction to T_1)	9	16
50% Reduction in Clearance Time (17% Overall Reduction in T_1)	19	44
20 Minute Earlier Demand Reduction	32	0*
10% Demand Reduction after 40 Minutes	14	0*
20% Demand Reduction after 40 Minutes	25	0*

*The off-peak traffic volumes of this example are not sufficient to be affected by these demand reduction situations.

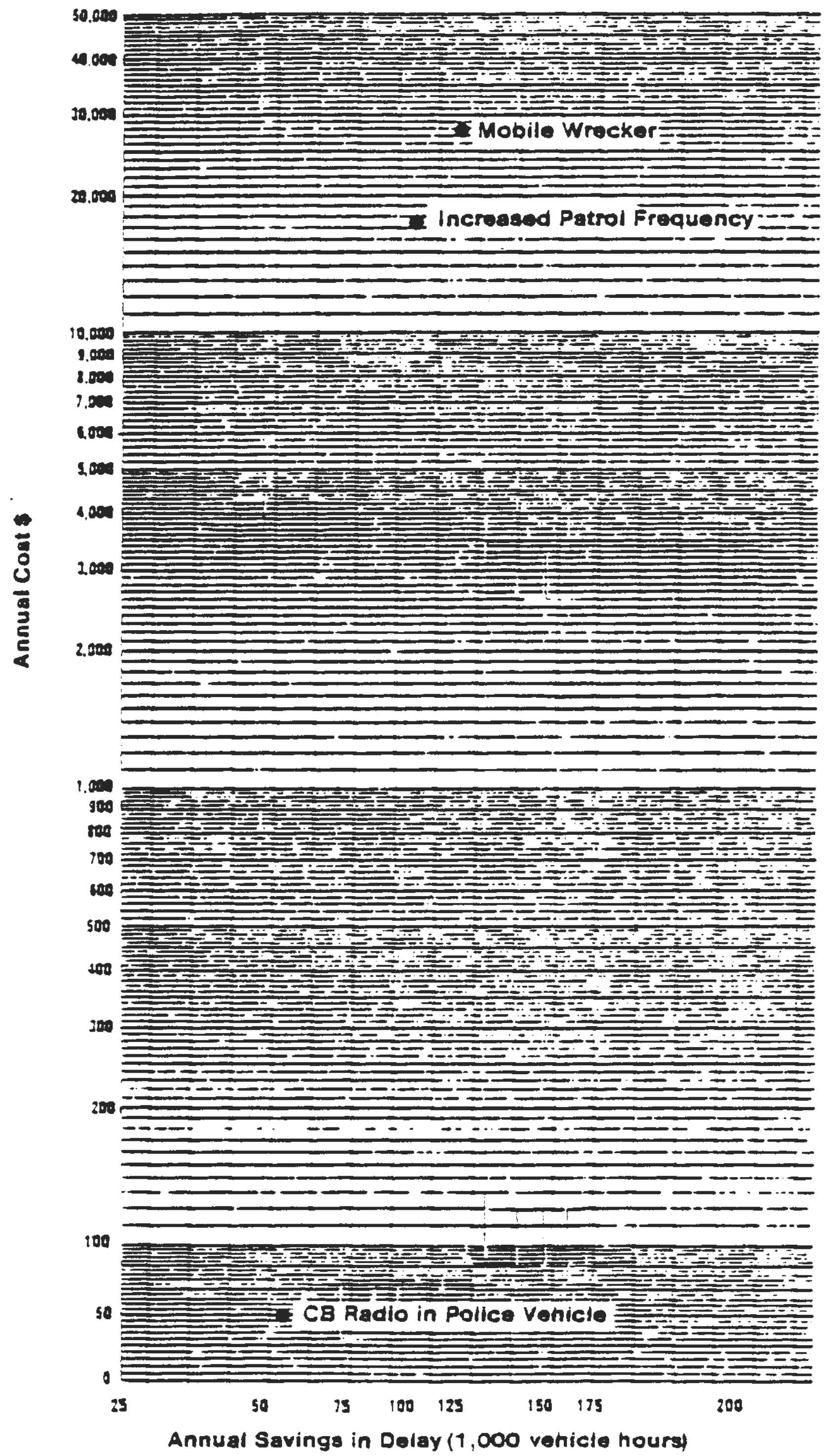


Figure 3.17. Cost-Effectiveness Plot

REFERENCES

1. Peat, Marwick, Mitchell & Company. Alternative Surveillance Concepts and Methods for Freeway Incident Management. Prepared for FHWA, RD-77-58 through 63 inclusive.

2. Pogust, F., Kuprijanow, A., and Forster, H., "Means of Locating and Communicating with Disabled Vehicles," National Cooperative Highway Research Program Report No. 6, 1964.

3. Evans, R. H., and Lam, T.N. "Gasoline Consumption in Urban Traffic," Research Labs, General Motors Corporation, Society of Automotive Engineers, Automobile Engineering Congress and Exposition (February 23 through 27, 1976).

This chapter discusses those impacts on the environment that result from traffic congestion. These are principally noise and air pollution.

NOISE

The Noise Problem

Noise is generally recognized as the major enviornmental problem caused by traffic in urban areas(1). Studies of noise in urban areas have indicated that traffic noise is usually the loudest and most frequently heard noise, and is the predominant traffic factor that disturbs people. Traffic noise becomes a particularly annoying factor when it penetrates the interior of buildings and interferes with conversation, sleep, and the enjoyment of radio and television. High noise levels can also often reduce the enjoyment of outdoor activities such as walking, shopping, and recreational pursuits.

The construction of urban freeways to relieve traffic congestion and remove through traffic from secondary and local roads often results in a reduction of the general levels of traffic noise in these latter areas. The concentration of this traffic and any additional traffic on freeways will, of course, create much higher noise levels in the immediate vicinity of the freeways. However, careful planning and design of the freeway corridor can protect the inhabitants of nearby buildings as well as people in nearby outdoor spaces from excessive noise levels. Before considering the ways in which this can be achieved, it is necessary to review the nature of traffic noise and the various standards that exist to measure noise and the nuisance it can cause.

The noise from individual vehicles emanates principally from the engine and exhaust system, and from the tires. For most private vehicles, these components of noise are aproximately equal under normal running conditions and they increase with speed. On acceleration, however, engine and exhaust noises tend to overshadow all others. Owing to their larger engines and heavier loads, trucks generally generate significantly higher noise levels than private cars. General traffic noise is complex and variable, and it depends on the type and distribution of vehicles using a road, the conditions of movement (whether vehicles are traveling at constant speeds, accelerating or decelerating, ascending or descending; see figure 4.1), the total volume of traffic, and the type of road surface.

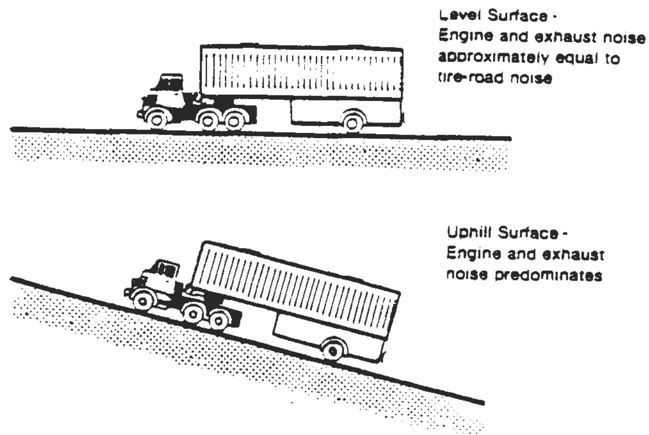


Figure 4.1. The Components of Traffic Noise

Traffic Noise on Urban Freeways

Traffic volumes are the main factor that affects noise levels on urban freeways. Noise increases in proportion to increases in traffic flow up to a certain point, but beyond which further increases in flow have little effect. The maximum noise level attained depends to some extent on the physical characteristics of the particular type of freeway. Typical curbside noise levels (which may be expected to be exceeded for 10 percent of the time) for flows in excess of 4,000 vehicles per hour, such as will occur on dual three- or four-lane freeways, will be in the region of 80 to 90 dBA (decibel A scale) during the day, and in the region of 70 to 80 dBA during the late evening. There is an area of uncertainty in the quoting of specific noise levels. Any values are highly dependent upon the precise conditions under which the readings were taken. There is a considerable range of values contained in the literature. Where applicable, we have quoted the limits of the values.

Traffic Noise in the Vicinity of Freeways

Traffic noise levels in the vicinity of urban freeways will obviously be less than those at the source of the freeway itself. The amount of reduction depends on the type of freeway (whether it is elevated, at ground level, or depressed), the distance from source to receiver, the the degree of shielding presented by topographical features, such as buildings and barriers. Figure 4.2 indicates the noise levels that can be expected at different types of freeway topography. However, the great variety of possible situations indicates that noise generation must be carefully considered for each individual case.

The noise levels to be expected at ground level in the immediate vicinity of depressed freeways with retaining walls will be less than for depressed freeways in open cutting. Both types of depressed freeways provide more protection than an elevated or ground-level freeway, unless these have acoustic barriers along their entire length. Farther away from the freeway, the reduction in the amount of noise experienced at ground level is also generally greatest in the case of the depressed freeway. However, with the addition of acoustic barriers along an elevated freeway, the reduction in noise levels will be marginally greater than for depressed freeways with or without barriers. Noise levels directly beneath an elevated freeway will be less than those at ground level elsewhere within an area of about 50 feet on either side of the freeway.

Similar comparisons can be made for the noise levels to be expected above ground in the vicinity of freeways. In general, unless very high barriers are erected along elevated sections, depressed freeways will provide more noise protection than elevated freeways.

The noise levels inside buildings in the vicinity of freeways will generally be less than noise levels outside. The amount of reduction depends on the type of materials used for the external walls of a building and whether windows are open. For buildings with open conventional windows, an outside to inside reduction of 5 to 19 dBA can be expected. Buildings, such as offices, which have closed, double glazed windows and mechanical ventilation have a reduction of around 30 to 35 dBA.

It is very difficult to relate particular noise levels to the nuisance they may cause because of the complex and completely random nature of traffic noise over any typical 24-hour period. The variation in individual tolerance to noise further complicates the task of assessing its nuisance.

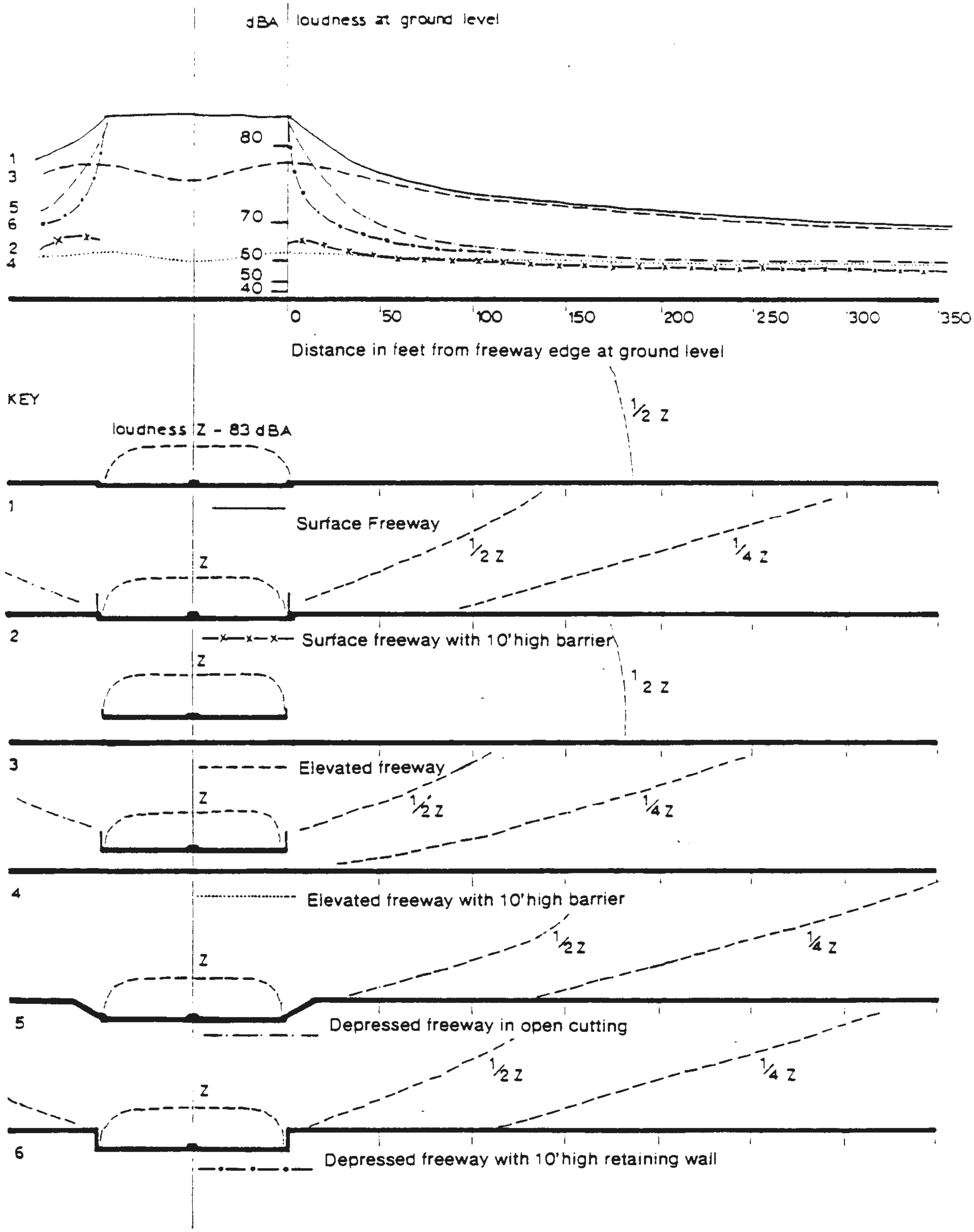


Figure 4.2. Noise Levels in the Vicinity of Various Types of Urban Freeway

Home interview studies in the United States have shown that the annoyance expressed with nuisance from noise depends not only on the levels of traffic noise themselves, but also on socioeconomic factors and general attitudes toward motor vehicles and urban freeways(2). These in turn may reflect the visual impact of the freeway and of other environmental factors. For example, in a particular area it was found that people in a high socioeconomic group were less tolerant to noise than those in a low socioeconomic group, even though the latter were more exposed to noise. The studies concluded that reactions were mixed and that a prediction of reactions must include consideration of psychological, as well as physical, factors.

Noise Measurement

Noise intrinsically varies over very short time periods. It covers a broad range of frequencies and its range of amplitude is extremely wide. These factors make noise measurement difficult, particularly as they must be related to the perception of the listener and thus, the response of the human ear must be taken into account. If noise is defined as unwanted sound, it will vary between individuals and situations.

In general, the average human ear can respond to frequencies between 20 to 20,000 Hz (cycles per second). However, not all frequencies in this range are heard equally as well as others. The frequencies between 500 and 5,000 Hz are more easily detected by the human ear. There is also a condition of the human ear that masks low-frequency sounds by a band of noise that is over one and a half octaves greater than low-level noise. These conditions, as well as others, resulted in establishing the dBA scale. Traffic noise measurements are made in this dB(A) scale. The commonly used values of L_{10}, L_{50}, and L_{90} refer to those sound levels that are exceeded 10, 50, and 90 percent of the time. Table 4.1 shows example values of noise levels of differing intensitites(3).

The Control of Noise

Some control on the amount of noise generated by freeways, and on the extent of the area in which it can be heard, can be exercised when freeways are planned. This can be done by avoiding the use of steep gradients, by locating access ramps and slip roads near intersections, and by the use of depressed freeways, particularly cut and covered sections, wherever possible. However, in many cases the choice of whether to use an elevated or a depressed road will also be influenced by other factors such as traffic movement, visual implications, severance, and engineering and economic considerations for the particular area concerned.

The most effective method of reducing noise levels in areas beside a freeway is to use some form of barrier such as a building, wall, screen, or landscaped earth mound. Barriers should be constructed close to, or integrally with, the road depending on the degree of noise reduction required and on the extent to which they are visually and economically acceptable. These barriers will reduce the noise level inside buildings and the external noise level. This makes it possible for people to open their windows without being exposed to intolerable noise. The degree to which an acoustic barrier shields noise from freeway traffic depends on the effective height of the barrier and its distance from the source of noise.

The greatest shielding is obtained with barriers as close as possible to the source of noise. Barriers midway between the source and the receiver provide the least noise reduction. When a barrier near the source is not possible, or when shielding for a particular point near the freeway is

TABLE 4.1. EFFECTS OF NOISE AND EXAMPLES OF NOISE LEVELS

	Noise Effects	Decibels	Typical Examples
Damage	Blast deafness	150	Explosions
	Pain	140	Engine tests
	Threshold of feeling	120	
			Thunder, gunfire
		110	Pneumatic drill
			Airplane
		100	
Annoyance	Reduced working efficiency		Subway
		90	Busy Street
	Occupational deafness	85	
			Noisy factory
		80	
	Interference with normal speech		Noisy office
		70	Suburban train
		65	
	Annoyance		Factory
		60	
Acceptable background levels			Large shop
		50	Quiet office
			Average house
	Comfort	45	
		40	
			Country road
		30	Quiet conversation
			Whisper
		20	
			Quiet church
		10	Soundproof room
			Threshold of hearing

TABLE 4.2. AVERAGE NOISE REDUCTIONS OBTAINED BY VARIOUS MEASURES

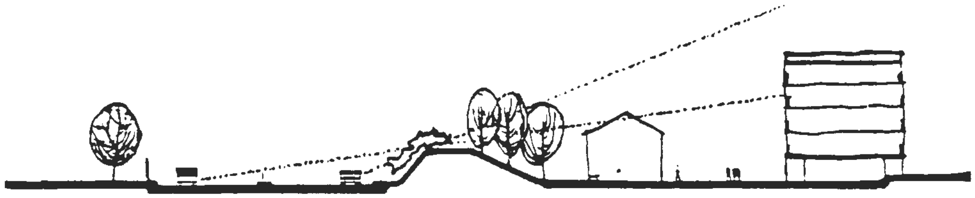
Method of Noise Reduction	Average Noise Reduction—dB(A)
Distance	3-6 per doubling of distance
Solid barriers 10 feet high on both sides of grade road	4 to 6
Road in 15-foot retained cutting	5 to 7
Road in 15-foot cutting with 10-foot high barriers	10 to 12
Cut and cover tunnel	35 to 45
100-foot thick belt of dense foliated trees	1 to 3
Open window	5 to 19
Closed conventional window	10 to 15
Double glazed windows with 8-inch spacing and mechanical ventilation	35 to 45
9-inch thick brick wall	45 to 55

required, it is best to position the barrier as near the receiver as possible. The degree of noise reduction is difficult to estimate precisely because of the variables of effective height and distance, but on average it can be expected to be between 6 and 15 dBA, and in some cases up to 20 dBA.

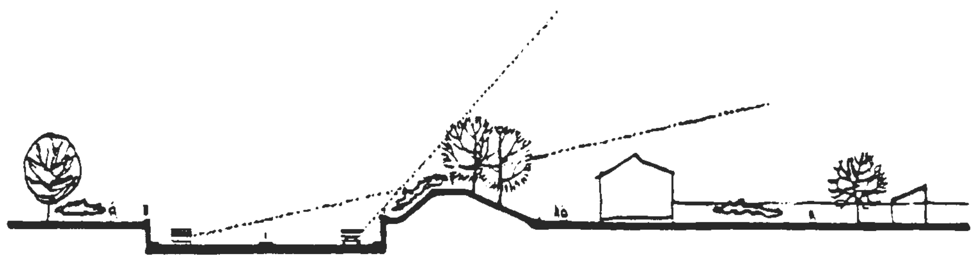
Table 4.2 shows some examples of the average noise reduction obtained by various countermeasures.

The structural form of barriers designed as part of freeway stuctures should be visually acceptable both to the pedestrian and the motorist. High barriers along an elevated freeway may increase the visual dominance of the structure on its surroundings. If the barrier is significantly higher than the motorist's eye level, it may be necessary to provide regular gaps in the barrier to prevent boredom and to allow a certain degree of directional guidance from glimpses of the surrounding landscape.

The use of landscaped earth mounds as noise shields is often possible for ground level or depressed freeways, particularly in outer urban areas or where freeways are set in parkland areas (figure 4.3). Widening the cuttings of depressed freeways can often reduce the height of retaining walls, consequently reducing the cost of construction and creating more pleasant views for the driver. The reduction of noise by trees and shrubs alone depends upon the type and density of these landscaping elements.



Surface Freeway



Depressed Freeway

The use of landscaped embankments as a noise shield is often possible, particularly in outer urban areas or where freeways are in parkland settings.

Figure 4.3. The Use of Landscaped Embankments

POLLUTION

The Problems

Air pollution from traffic fumes is of concern in the United States where, in places such as Los Angeles, climatic conditions can be responsible for converting certain components of these fumes into an unpleasant smog.

Although the concentration of traffic onto freeways will increase air pollution in the freeway corridors, the effects will be extremely local and are likely to be a minor problem for the occupants of nearby developments when compared with the nuisance of traffic noise and loss of privacy.

The major source of fumes from motor vehicles is their exhaust systems. The fumes emitted contain carbon monoxide and unburned hydrocarbons, which are the result of incomplete combustion, together with small quantities of carbon dust particles, nitrogen oxides and lead compounds. Other hydrocarbon fumes are emitted from carburetors, gas tanks, and crankcases. Hydrocarbon fumes produce a smell that can cause annoyance, and can produce smog in those areas where high temperatures and strong sunlight cause them to combine with nitrogen oxides.

TABLE 4.3. EMISSION RATES
(Grams Per Minute)

	Hydrocarbons (HC)	Carbon Monoxide (CO)	Nitrogen Oxides (NO_x)
Idle	1.5	17.0	0.1
Free flow	1.4	12.4	0.5
Acceleration/Deceleration	1.7	19.4	1.1
Incident Queueing	1.6	17.6	0.4

The amount of engine and exhaust fumes generated on urban freeways depends on traffic volumes and the amount of fumes generated by individual vehicles. This, in turn, depends on the engine capacity, running speeds, and the efficiency of fuel combustion of each vehicle. Because they are heavier than air, exhaust fumes tend to hang low in the air in the region where they are generated. The build-up of fumes will depend on the type of freeway and the extent to which it reduces or encourages the free circulation of fresh air.

On elevated sections of urban freeways, the build-up of fumes is prevented by the constant agitation of the air by traffic, therefore, conditions are unlikely to become unpleasant for drivers. Fumes can collect immediately alongside or underneath elevated structures, particularly if they are augmented by the fumes from traffic on access ramps or secondary roads adjacent to or underneath the structure. In such cases, the support columns and other vertical surfaces of abutments and ramp walls may, in time, become dirty as a result of the deposit of exhaust particles. Unlike the unsightly dirt stains that can occur as a result of splash and dirt thrown up from the tires of passing vehicles, dirt deposits from fumes are more uniform in appearance.

Traffic fumes are more confined to the limits of depressed freeways, particularly long covered sections or tunneled sections, which consequently accumulate higher concentrations than elevated sections. Unsightly deposits of exhaust particles adhere to the retaining walls of depressed freeways and driving conditions become unpleasant. This may arise as a result of exhaust fumes entering vehicles through low-level air intakes or windows. In most cases, these fumes will only be a minor nuisance to drivers, but if the build-up of carbon monoxide becomes excessive, it could result in drowsiness and subsequent loss of judgment. While this could be a serious hazard, there is little evidence to suggest that it has been the cause of accidents.

Table 4.3 shows the emissison rates in grams per minute for the principal pollutants emitted by vehicles(4). The values for the various stages of queueing, accelerating, etc., can be used in conjunction with the delay computation equations in chapter 3 of this volume to estimate likely changes in pollutants produced as a result of alternative incident or congestion management strategies.

REFERENCES

1. British Road Federation. "Monitoring in the Urban Environment." London, 1971.

2. National Cooperative Highway Research Program. "Highway Noise Measurement Simulation and Mixed Reactions." NCHRP Report No. 78, 1969.

3. Hobbs, F. D. Traffic Planning and Engineering, University of Birmingham, England. 1979

4. FHWA. Alternative Surveillance Concepts and Methods for Freeway Incident Management. Prepared for FHWA-RD-77-59. 1978.

CHAPTER 5. USER SIMULATION MODELS

Computer simulation models are a tool that the engineer can use to evaluate alternative designs and control strategies for a freeway or a freeway corridor. The usefulness and practicality of simulation models has increased greatly in the past two decades as digital computers have decreased in cost and increased in capacity and efficiency. Computer simulation provides the means for controlled experiments. Each strategy systematically varies traffic conditions, evaluates roadway performance, and determines the point at which the facility has been overloaded. Such detailed information could not be obtained by field tests, even if the inconvenience to the public could be tolerated(1). In many cases (particularly large-scale projects where numerous alternatives are being considered) simulation is the only feasible method of estimating the impacts of different alternatives.

The geometric representation of a roadway system for most freeway models is comprised of links (one-directional roadway segments) and nodes (intersections or geometric discontinuities). The logical division of a road system into links may correspond to the natural segmentation caused by cross streets on ramp junctions. If analysis of a natural segment indicates different characteristics on one portion rather than on another, it may be desirable to further subdivide the segment. A change in grade is an example of a reason to subdivide a segment. Links are also classified as surface, ramp, or freeway links, to account for the different operating characteristics of these facilities.

Numerous freeway simulation models have been developed at this time. Most of the models were created for a specific purpose or type of operation. In 1974, Hsu and Munjal(2) reviewed 15 freeway simulation models associated with various aspects of freeway vehicular traffic and the models are compared against a baseline of eight desirable model features. The reader may wish to refer to the paper by Hsu and Munjal for information on these earlier models:

1. Arizona Transportation and Traffic Institute Traffic Simulation Model(3)
2. Midwest Research Institute Freeway Simulation Model(4,5)
3. Midwest Research Institute Mountainous Terrain Model(6)
4. Northwestern University Lane-Changing Model(7)
5. Sinha Freeway Simulation Model(8,9)
6. Connecticut Department of Transportation Expressway Simulation Model(10)
7. Texas Transportation Institute Freeway Merging Model(11)
8. System Development Corporation Diamond Interchange Model(12)
9. System Development Corporation Freeway Simulation Model(13)
10. Mikhalkin Freeway Simulation Model(14)
11. Georgia Model(15)
12. SCOT Corridor Model(16)
13. Priority Lane Model(17,18)

14. Aggregate Variable Models(19)

15. Aerospace Corporation Freeway Simulation Model(20)

Many of these models were early versions or parts of current models. For example, Lieberman and Bullen used Model 4 in the development of the INTRAS model while Model 12 was the early version in the SCOT model family. Model 13 was an early priority lane version in the FREQ model family, and Model 14 was the early version in the MACK model family(21).

There are currently five families of freeway corridor simulation models available. The five families are CORQ, FREQ, INTRAS, MACK, and SCOT.

CORQ-CORCON MODEL

The CORQ model developed by Yagar during 1968 to 1975 and the related CORCON model developed by Allen and Associates during 1974 to 1978 are the two models in this family. In addition to their development, these models have been applied in Ottawa, San Francisco, and Toronto by the developers(21).

CORQ, written in FORTRAN IV, gives detailed treatment of the critical elements of a corridor in terms of traffic flow, capacity, queuing, and delays. It is related to another model, FREQ, which emphasizes the modeling of freeway queues(22). CORQ is a form of micro-assignment technique, but it is different from most of the existing techniques. For instance, it is completely different from the Brown and Scott technique(23) although both can be used for micro-analysis of areas that are about 500 blocks large. The methods accomplish this by totally different micromodeling procedures. The Brown and Scott model considers all intersections, but CORQ handles only major intersections, the freeway interchanges, and freeway and surface-street links between them. However, it gives a more detailed treatment, especially to the intersections. It also can be used for much larger areas if, for example, only the freeway network needs to be modeled. Another major difference is that CORQ treats all time-varying demands, and the Brown and Scott model seems to treat only homogeneous demand tables with a constant O/D pattern although it does allow the rate of demand to vary with time(24).

The basic method of CORQ is to divide the peak period into a set of sufficiently short time slices of common length so that the rates of demand between the various O/D pairs can be considered constant for about 15 minutes. This allows the time-varying demand to be expressed as a set of O/D matrices representing the respective time slices; each slice has stationary demands. The O/D matrices are assigned to the network sequentially in time. This allows temporary oversaturation of network links. That is, in any time slice, certain network links may have more demand assigned to them than they can serve. Excess vehicles queue on upstream links and are reassigned to their destinations in the succeeding time slices from the points at which they queued. The assignment is based on the principle of minimum individual travel cost, and the minimum cost path may include some time in queue(24).

CORQ has the capability for changing network characteristics at the beginning of each time slice because capacity variations may be as important as demand variations (for example, those that simulate transient traffic controls such as time-varying, ramp-metering rates).

The following outlines the logic of the model:

1. Routine for each time slice:

a. Note any changes in network characteristics that take place in a time slice.

b. Set O/D matrix equal to demand for the new time slice plus any queues from previous time slice.

2. Routine for each incremental assignment of the iteration:

   a. For each origin node, O_i, having some demand, find tree of shortest paths to all destinations.

   b. For each destination node, D_j, work back to the origin, and note the first point of congestion in the O/D path.

   c. For each destination node, D_j, tentatively assign those flows and queues that would result if all the remaining demand from O_i to D_j were assigned.

   d. Find the critical sublink that limits the fraction of the tentatively assigned flows and queues that actually can be assigned in that increment.

   e. Assign the appropriate fraction of the tentative assignment as determined by the critical limiting link.

   f. Estimate the weave section capacities on the basis of the assigned flows.

   g. If it is desired to dynamically share the merge capacity, estimate the component capacities for each merge on the basis of weave capacity, respective merge entitlements, and assigned merge flows.

   h. Update the statistics for each link.

   i. If the entire O/D matrix has not been assigned, perform the incremental assignment routine again(24).

The CORQ model was tested on the Ottawa Queensway corridor(25). The flows and queues that it initially predicted were reasonably close to those measured in the field. Therefore, it was calibrated to actual flows and queues, and applied in testing alternative traffic-control schemes(26). It was further validated in application, where it demonstrated its sensitivity in modeling the effects of various strategies and its power in suggesting alternative paths for some bottleneck users.

The data requirements for CORQ are of the type generally collected when doing a freeway control study. These include capacities, counts, queue sizes, travel-times as a function of flow and origin/destination (O/D) information on users who could or should be affected by controls. Unless these data are already available, the task of collecting the data solely for use in the model would be prohibitive.

The CORCON model developed by Allen and Easa(27-31) is an analytical procedure for predicting traffic volumes, queuing conditions, and travel times in a freeway corridor. Traffic demand can vary over time and is assigned to a freeway and surrounding arterial street network. The minimum path algorithm can incorporate turn prohibitions. A major characteristic of the CORCON model is its link-node representation that simplifies network representation by allowing more than one directional roadway link to have common upstream and downstream nodes and automatically avoid illogical paths in the network(21).

The data requirements for CORCON are similar to those of CORQ. Origin/destination demands, link volumes and queues, flow versus travel time relationships, capacity information, and network turn prohibitions were collected for using CORCON on a study in Toronto by Allen et al(31). In the Toronto study, CORCON was used for predicting traffic-operating characteristics in the corridor network before and after entry control. The model can be used not only to evaluate the effects of freeway ramp metering but also to evaluate the impact of a wide range of Transportation System Management strategies(31).

Because of dimensioning constraints in the currently available computer program, it may be difficult to apply CORCON to very large corridors (5 x 25 miles). This can be accommodated by either increasing the dimension sizes, analyzing two or three smaller subsections of the corridors separately, reducing the amount of coded network detail, or some combination of the preceding.

FREQ MODEL FAMILY

Work on the first model of the FREQ family of models began in 1968 at the University of California, Berkeley. From this first model, FREQ(32), extensions and refinements were made to produce FREQ2 and FREQ3 with particular attention being directed to shock wave analysis, computer efficiency, and output format(33,34).

In the early 1970's, FREQ3CP, FREQ3D, and FREQ3C were developed and incorporated priority entry control, design improvement, and normal entry control optimization sub-models, respectively (35, 36, 37).

The latest versions in use at this time are the FREQ6PE and FREQ6PL models (38, 39). The FREQ6PE model is a macroscopic decision model of a freeway corridor and is used primarily for the evaluation of priority entry and normal entry control on a directional freeway. The model can also be used for evaluating design improvements with or without freeway entry control. The model predicts a time stream of impacts and traveler responses due to the interaction between ramp control strategy and traveler responses. The impact assessment includes traveltime, fuel, emissions, and noise while demand forecasting includes spatial and modal traveler responses in increments during the first year of operation.

The FREQ6PL model is a macroscopic model of a freeway corridor, used primarily for the evaluation of reserving lane(s) on freeways for carpools and/or buses. The model can also be used for evaluating design improvements with or without priority operation. The user selects the priority lane(s), design configuration, priority cut-off level, and time duration of priority operations. The model automatically modifies the demand and supply sides of the model, and predicts a time stream of impacts and traveler responses. The impact assessment includes traveltime, fuel, emissions, and facility costs while the demand forecasting includes spatial and modal traveler responses in increments during the first year of operation.

The FREQ family of models is written in FORTRAN IV and can be operated on IBM and CDC computer systems. As with the other models, the data requirements to operate the FREQ models are quite extensive. The freeway is broken into homogeneous subsections that are usually stretches of freeway between on- and off-ramps, changes in freeway width, etc. The physical characteristics of the freeway necessary for input are:

- Subsection length

- Number of lanes in subsection
- Lane width in subsection
- Subsection capacity
- Subsection truck factor
- Subsection grading
- Length of grade for subsection
- Subsection design speed

In addition to the physical characteristics of freeway subsections, the model considers ramp characteristics:

- Location of on- and off-ramps
- Characteristics of special ramps (multi-lane ramps, left-hand side ramps, etc.)
- Ramp metering limit and/or capacities

The FREQ models also require demand data, which consists of:

- An origin/destination table for each time slice—this may be in the form of traffic counts or hourly rates and may be either measured or simulated.
- Passenger volume data—in some cases, it is desirable to measure passenger throughput (as in priority entry control simulation). The user may vary the average vehicular occupancy during each time slice.

For the FREQ6PE model, data are also required on the arterial capacity and flow rates, as well as a measure of the progression along the arterial.

According to the FREQ6PE users' manual(38), the cost for simulating 30 subsections at the lowest priority on a CDC 6400 computer is approximately $25.00.

INTRAS

INTRAS was developed for use in studying freeway incident detection and control strategies(40). It is based on knowledge of freeway operations and surveillance systems, and incorporates detailed traffic simulation logic developed and validated for use in the model.

To allow simulation of freeway control policies, including ramp metering and diversion, the capability of modeling the off-freeway environment is included in INTRAS. This "surface" traffic modeling is patterned after the logic of the UTCS-1 simulation model(41).

To facilitate the simulation of closed loop incident detection and control, as well as off-line traffic analysis, the INTRAS model contains a realistic surveillance system simulation capability. The ability to visualize vehicle trajectories, and contours of measures of effectiveness (MOE's) in the time-space plane, is included in INTRAS via a digital plotting module. INTRAS also contains a statistical analyses module that permits comparison of MOE's from different simulation runs or field data, utilizing standard parametric and non-parametric tests.

Finally, a fuel consumption and vehicle emission evaluation module is built into INTRAS patterned after a similar module developed for the UTCS-1 simulation model.

INTRAS allows the user to simulate an incident at any location on a freeway link and for any length of time. The incident may block one or more lanes or be confined to the shoulder.

Three types of traffic detectors can be simulated by INTRAS: Doppler radar detectors, short inductance loops, and coupled short inductance loops.

At each freeway on-ramp, the user may simulate one of four types of ramp control. The four types are: clock time metering, demand/capacity metering, speed control metering, and gap acceptance merge control.

Data required for input to INTRAS include geometric data to describe the links, such as: link length, number of lanes, lane channelization, type of link, grade, radius of curvature (for freeway links) percent of superelevation, and pavement type. Operational data required for input include: entry link flow rates, percentages of intercity buses, heavy single-unit trucks, trailer trucks, high-performance passenger cars, low-performance passenger cars, turning percentages at intersections, discharge headways, lost time, and free-flow speeds. Also required for input to INTRAS are control data to identify the type of control at intersections; i.e., stop sign, fixed time or actuated signal, the actual signal timing, ramp control operation, location and type of detectors present, and location, type, and time of incident.

The INTRAS model is written in standard ANSI FORTRAN. The model was developed for and implemented on the CDC 6600/7600 computer series and the IBM 360/370 computer series. Since ANSI FORTRAN is used, the model may be implemented on virtually any large-scale computer having an ANSI FORTRAN compiler.

MACK FAMILY

The MACK model and its later versions are deterministic, macroscopic models that consist basically of a set of conservation equations and a corresponding set of dynamic speed-density equations. Payne and associates began work in the late 1960's, and models in this family include MACK I, MACK II, MACK III, and FREFLO. In the MACK I model, the dynamics of traffic are described by the numerical solution of fluid-flow differential equations, appropriately modified to represent the traffic environment. This model generates results that exhibit all the global dynamic responses to freeway traffic flow, with a minimum of computational costs. It is capable of simulating the response of the system to incidents that block one or more lanes on a section of roadway. The model does not distinguish flow by lanes.

In the MACK II model, a new equilibrium speed-density relationship and a structural change in the dynamic speed relationship involving the parameters were introduced. The MACK II and INTRAS models were applied to a segment of the Shirley Highway outside of Washington, D.C., with incident-free and incident scenarios. The authors recommended that MACK II be adopted for the purpose of making preliminary evaluations of control strategies for responding to incidents based on the qualitative agreement between MACK II and INTRAS results, and the 100-fold cost factor between execution of the macroscopic MACK model and the microscopic INTRAS model(42,43).

The FREFLO model was developed by Payne at ESSCOR and was a successor to the MACK II model. The FREFLO model was designed to provide a basic model, perform input data diagnostics, represent incidents, model on-ramps, control time of day, represent surveillance, represent two traffic-responsive metering schemes, provide standard measures of travel and travel time, include fuel consumption, and include pollution emissions.

The geometric data required by FREFLO consist of the number of lanes in each section, section lengths, on-ramp and off-ramp locations, and nominal section capacities. The traffic data required are densities and speeds of each section for the initial state, upstream freeway volume for each time period, and on- and off-ramp volumes for each time period.

SCOT

The SCOT model is a combination of two simulation models: DAFT and UTCS-1 (44). It has the capability of simulating traffic patterns within an integrated freeway corridor. It is macroscopic on the freeway elements, microscopic on the ramps, service roads, major arterials, and city streets. Traffic performance measures may be computed for a wide variety of control strategies; e.g., traffic signal schedules, ramp metering policies, and bus priority schemes. The effect of these strategies on traffic flow in each element of the corridor can now be determined quantitatively; e.g., the increase in freeway average speed, the growth of queues on the ramps, the number of vehicles moving through an intersection, and the location, extent, and duration of congestion.

The DAFT model is a macroscopic simulation of traffic along a network of freeways, ramps, and arterials. Vehicles are grouped into platoons and lose their individual characteristics. The platoons are moved along the freeway according to a single, prespecified speed-density relation that applies to all freeway links. Along the non-freeway links, they travel at the specified free-flow speed for each link and are delayed at the downstream end for a time related to the ratio of green time to signal cycle time (G/C) of the facing signal there and to the volume of traffic on that link. For each entry link at the periphery of the network considered, traffic volumes are specified according to destination node; i.e., the input data consist of origin/destination (O/D) demands that vary with time. The model distributes the resulting platoons of vehicles over the network so as to satisfy these specified O/D patterns and to follow minimum-cost paths. These minimum-cost paths are calculated frequently by the model, based on current conditions. Whenever a platoon reaches a network node, its turning movement there is dictated by its minimum-cost path as it exists at that instant of time. Hence, the model produces a dynamic assignment of traffic as a by-product of the simulation.

The basic approach adopted in the design of the SCOT model was to combine the DAFT and UTCS-1 models appropriately. The microscopic logic of UTCS-1 is applied to those components of the network characterized by signalized, grade intersections. Here, the traffic mechanisms are so complex, because of the many conflicts common to the stop-and-go patterns of urban traffic, that each vehicle must be treated individually, and a very small time step must be applied to obtain an acceptable level of accuracy in the replication of global traffic flow. Traffic along the freeway, however, has been modeled successfully by the use of fluid-flow analogies and has often been referred to as stream flow. It was believed that applying a microscopic approach to freeway traffic would yield little in the way of additional global accuracy. Furthermore, a microscopic treatment of freeway traffic would greatly magnify computing costs and storage requirements.

The SCOT output is a dynamic history of the state of traffic on each street and intersection during the simulation period. In addition, averages for the entire network are computed. Thus, a local and a systemwide view is obtained.

The locations and durations of spillback (the extension of a queue into the upstream intersection blocking cross traffic) and the instances of cycle failure (the inadequacy of a green interval duration to discharge the entire queue) are recorded. The number of vehicles currently on each street is given. Cumulative statistical data for each street as well as the network as a whole are also available as often as desired. These include the number of vehicles that have traversed each street, the number of stops made on each street, average speed, average delay per vehicle and mean occupancy on each street and for the network as a whole, vehicle-miles, vehicle-minutes, delay per vehicle-mile, and traveltime per vehicle-mile. Bus statistics are compiled separately and include the number of buses processed and the number of stops made on each street. Bus route data, such as average speed, total bus station dwell time, and total delay time for each bus route, are given.

The information required by SCOT to model a network is generally available to traffic engineers. These data fall into four sets: geometric, traffic demand, control system, and bus schedules.

Typical geometric data needed are spans of each street or freeway section, grades and spans of ramps and turning lanes. The number and widths of travel and parking lanes must be known as well as the locations of internal traffic generators such as shopping centers and parking lots.

Data that describe the volume and character of traffic must be obtained. Speed-density data are needed for each freeway section. In-flow rates at the periphery of the network as well as traffic mix, i.e., ratio of number of passenger cars to trucks, the usual turning movements and pedestrian activity at each intersection, and the street free-flow speeds and rate of queue discharge, are required.

All parameters of the control strategy are needed. These parameters include: traffic signal timings and synchronizations; parking restrictions and lane use (e.g., right turn only); ramp metering and bus priority policies; and traffic-actuated logic.

Bus service data needed include bus routes, station locations and capacity, frequency of bus service, and mean station dwell times.

Sperry researchers working on an FHWA-sponsored project concerned with the development of traffic logic for freeway corridor control developed a corridor optimization program called DIVSIM. This corridor optimization algorithm was embedded in the SCOT model(45).

A summary table of the major simulation models is presented in table 5.1.

TABLE 5.1. SUMMARY OF SIMULATION MODELS

Model and Type	Required Input Data	Output	Advantages	Disadvantages
CORQ Macroscopic	Capacities, volumes, queue sizes, travel-times as a function of flow and origin/destination (O/D) information	Traffic flow, capacity, queuing, and delays	Detailed treatment of major intersection and freeway interchanges; treats all time-varying demands	Considers only major intersections; extensive O/D information required
FREQ Macroscopic	Subsection characteristics such as: length, number of lanes, lane width, capacity, truck factor, grading, length of grade, design speed. ramp locations, ramp characteristics, metering limits, O/D table, passenger volume	Traveltime, fuel, emissions	Evaluation of priority entry and normal entry control with FREQ6PE; priority lane evaluation with FREQ6PL	Extensive data requirements
INTRAS Microscopic	Link length, number of lanes, lane channelization, type of link, grade, radius of curvature, superelevation, pavement type, flow rates, percent buses, trucks, turning percentages, discharge headway, lost time, and free-flow speed, type of intersection control, ramp control operation, detector location and location type, and time of incident	Travel, traveltime, delays, stops, fuel, emissions	Ability to simulate incidents and evaluate incident control strategies	Large data requirements; very expensive to run

Table 5.1, Continued

Model and Type		Required Input Data	Output	Advantages	Disadvantages
MACK	Macroscopic	Number of lanes, section lengths, ramp locations, section capacities, initial speed and densities, upstream volume, and ramp volumes	Travel, traveltime, fuel, emissions	Low cost to execute	Does not distinguish flow by lanes
SCOT	Macroscopic-for freeway; microscopic-for other than freeway	Section lengths, grades, number and width of travel and parking lanes, internal traffic generators, speed-density data, traffic mix, free-flow speed, signal timings, metering policies, and bus data	On each street: the number of stops, average speed, average delay, mean occupancy; for the network: vehicle-miles, vehicle-minutes, delay per vehicle-mile, and travel-time per vehicle-mile	Detailed analysis of each link	Costly to run

REFERENCES

1. Muzyka, A. "Urban/Freeway Traffic Control." Presented at IEEE Conference on Decision and Control, New Orleans, Louisiana, December 13-15, 1972.

2. Hsu, Y.S. and Munjal, P.K. "Freeway Digital Simulation Models." TRB Transportation Research Record 509, 1974, pp. 24-41.

3. Richard, R.M., Baker, R.L., and Sheldon, W.P. "Simulation of Traffic Flow as a Basis for Interchange Design." Final Report, Arizona Transportation and Traffic Institute, for Arizona Highway Dept. and FHWA. June 1965.

4. Kobett, D.R. A Digital Simulation Model of Freeway Traffic. Vol. I: "Model Description." Final Report, Midwest Research Institute, Contract CPR-11-3661 for FHWA, December 1968.

5. Kobett, D.R., and Levy, S.L. "Study of Expressway Traffic Flow Through Digital Simulation." Final Report, Midwest Research Institute, Contract CPR-11-0963 for FHWA, May 1965.

6. St. John, A.D., Kobett, D.R., Sommerville, D., and Colanz, W.D. Traffic Simulation for the Design of Uniform Service Roads in Mountainous Terrain, Vol. I-IV. Final Report, Midwest Research Institute, Contract CPR-11-5093 for FHWA, May 1970.

7. Worral, R.D., and Bullen, A.G.R. "Lane-Changing on Multi-Lane Highways." Final Report, Northwestern University, Contract CPR-11-5228 for FHWA, August 1969.

8. Sinha, K.C. "The Development of a Digital Simulator for the Analysis of Freeway Traffic Phenomena." University of Connecticut, PhD dissertation, 1969.

9. Sinha, K.C., and Dawson, R.F. "The Development and Validation of a Freeway Traffic Simulator." Highway Research Record 308, 1970, pp. 34-47.

10. Leland, S.D. "A General Traffic Flow Simulation Model for Freeway Operation." Final Report, Connecticut Dept. of Transportation, May 1970.

11. Buhr, J.H., Meserole, T.C., and Drew, D.R. "A Digital Simulation Program of a Section of Freeway with Entrance and Exit Ramps." Highway Research Record 230, 1968, pp. 15-31.

12. Nemeczky, J.A., and Widdice, B.D. "Development and Validation of a Digital Simulation Model for Research in Geometric Design and Control of Diamond Interchanges." Presented at Summer Simulation Conference, sponsored by ACM/SHARE/SCI, 1970.

13. Warnshuis, Paul. "Simulation of Traffic Flows Through Interchange Systems of Arbitrary Configuration." System Development Corp., Contract FH-11-7628 for FHWA, March 1972.

14. Mikhalkin, B. "Estimation of Roadway Behavior Using Occupancy Detectors." University of Southern California, PhD dissertation, Feb. 1971.

15. Wildermuth, Bruno R. "Effect of Lane Placement and Truck Traffic Freeway Flow Characteristics." Final Report, Wilbur Smith and Associates for Georgia Highway Dept., Project 6909, April 1971.

16. Wicks, David. Traffic Flow Simulation Study: The SCOT Model, Vol. I-IV. Final Report, General Applied Science Laboratories, Inc., Contract DOT-TSC-161, Feb. 1972.

17. Blankenhorn, R.C., and May, A.D. "FREQ2: A Revision of the "FREQ" Freeway Model." Special Report, Institute of Transportation and Traffic Engineering, University of California, Berkeley, June 1972.

18. May, A.D., Minister, R.D., Lew, L.P., and Ovaici, K. "A Computer Simulation Model for Evaluating Priority Operations on Freeways." Special Report, Institute of Transportation and Traffic Engineering, University of California, Berkeley, Contract FH-11-7704 for FHWA, June 1972.

19. Payne, H.J. "Aggregate Variable Models of Freeway Traffic." Depart. of Electrical Engineering, University of Southern California, NSF Grant GK-24520, June 1972.

20. Harju, G.W., Bush, L.R., Kramer, R.F., and Porges, H.S. "An Advanced Computer Concept for Freeway Traffic Flow Modeling." Presented at Summer Simulation Conference, June 1972.

21. May, A.D. "Freeway Corridor Analysis." Presented at Conference on the Application of Traffic Simulation Models, June 3-5, 1981, Williamsburg, Virginia.

22. Makigami, Y., and Woodie, W.L. "Freeway Travel Time Evaluation Technique." Highway Research Record 321, 1970, pp. 33-45.

23. Brown, G.H.J., and Scott, R.S. "Micro-Assignment: A New Tool for Small-Area Planning." Highway Research Record 322, 1970, pp. 149-161.

24. Yagar, S. "CORQ—A Model for Predicting Flows and Queues in a Road Corridor." TRB Transportation Research Record 533, 1975, pp. 77-87.

25. Yagar, S. "An Analysis of the Morning Peak Traffic Problems in the Eastbound Corridor Serving Ottawa." Ontario Ministry of Transportation and Communications, 1973.

26. Yagar, S. "Measures of the Sensitivity and Effectiveness of the CORQ Traffic Model." TRB Transportation Research Record 562, 1976, pp. 38-48.

27. Easa, S.M., and Allen, B.L. "Analysis of Urban Freeway Corridors: Toward the Development of a Model for Assessing Freeway Control Plans." McMaster University, Interim Report 3311-76/1, July 1976.

28. Allen, B.L., and Easa, S.M. "Analysis of Urban Freeway Corridors: A Corridor Model for Assessing Freeway Control Plans." McMaster University, Final Report 3311-72/2, Vol. 1, March 1977.

29. Allen, B.L., and Easa, S.M. "Analysis of Urban Freeway Corridors: Field Validation of CORCON—A Corridor Model for Assessing Freeway Control Plans." McMaster University, Final

Report 3311-77/3, Vol. 2, March 1977.

30. Easa, S.M., and Allen, B.L. "CORCON: Freeway CORridor Assignment and CONtrol Model." Paper presented at the 57th Annual Meeting, Transportation Research Board, 1978.

31. Allen, Brian L., Easa, Said M., and Case, E.R. "Application of Freeway-Corridor Assignment and Control Model." Transportation Research Record 682, 1978, pp. 76-84.

32. Makigami, Y., Woodie, L., and May, A.D. "Bay Area Freeway Operations Study." Final Report. Part I of III: The Freeway Model. Berkeley: University of California, Institute of Transportation and Traffic Engineering, August 1970.

33. Blankenhorn, R.C., and May, A.C. "Freeway Operations Study-Phase III." Interim Report No. 4-1: FREQ2-A Revision of the "FREQ" Freeway Model. Berkeley: University of California, Institute of Transportation and Traffic Engineering, June 1972.

34. Stock, W.A., Blankenhorn, R.C., and May, A.D. "Freeway Operations Study—Report III," Report 73-1: The FREQ3 Freeway Model. Berkeley: University of California, Institute of Transportation and Traffic Engineering, June 1973.

35. Ovaici, K., May, A.D., Teal, R.F., and Ray, J.K. "Simulation of Freeway Priorities Strategies (FREQ3CP)." Final Report, Federal Highway Administration, June 1975, 471 pages. (Also published in Proceedings of Sixth International Symposium on Transportation and Traffic Theory, Sydney, Australia, August 1974; TRB Record 533, 1975; and available as UCB Dissertation.)

36. Eldor, M., and May, A.D. "Freeway Operations Study—Phase III." Report 73-1. Cost Effectiveness Evaluation of Freeway Design Alternatives. Berkeley: University of California, Institute of Transportation and Traffic Engineering, June 1973. (Also published in Proceedings of Engineering Optimization Conference, Liverpool, England, August 1974; and in TRB Record 495, 1975.)

37. Wang, J.J., and May, A.D. "Freeway Operations Study—Phase III." Report 73-3. Analysis of Freeway On-Ramp Control Strategies. Berkeley: University of California, Institute of Transportation and Traffic Engineering, June 1973. (Also published in HRB Record 469, pp. 16-25, 1973; and is available as UCB Dissertation.)

38. Jovanis, P.P., May, A.D., and Yip, W. "FREQ6PE—A Freeway Priority Entry Control Simulation Model." California Department of Transportation. Final Report, September 1978, approximately 300 pages. (Vol. II is for limited distribution and contains computer-related materials.)

39. Cilliers, T., May, A.D., and Cooper, R. "FREQ6PL—A Freeway Priority Lane Simulation Model." California Department of Transportation, Final Report, September 1978, approximately 300 pages. (Vol. II is for limited distribution and contains computer-related materials.) (Presented to TRB for possible presentation and publication, January 1979. Also available as UCB Dissertation.)

40. Wicks, D.A., and Andrews, B.J. "Development and Testing of INTRAS, A Microscopic Freeway

Simulation Model," Vol. 2, Users Manual. Final Report, FHWA-RD-80-107, July 1980.

41. Peat, Marwick, Mitchell, & Co. and KLD Associates, Inc. "Network Flow Simulation for Urban Traffic Control System-Phase II," Report No. DOT-FH-11-7885, 1973.

42. Payne, H.J., and Young, D.C. MACK: A Macroscopic Simulation Model of Freeway Traffic (Version II). Orincon Corp. April 1977.

43. Payne, H.J., and Koble, H.M. A Comparison of Macroscopic and Microscopic Simulation Models (MACK and INTRAS). Orincon Corp., Report OC-R-78-11-9367-1, March 31, 1978.

44. Lieberman, E.B., and Wicks, D.A. "Traffic Flow Simulation Study—The SCOT Model." Highway Research Record 409, Washington, D.C., 1972, pp. 34-45.

45. Sperry Systems Management. Optimizing Traffic Flow in an Intercity Corridor. Vol. I: Program Discussion and User Guide." Vol. II: "Sub-Routine Description." Vol. III: "SCOT Model Modification and Traffic Control." Vol. IV: "Derivation of Traffic Logic." FHWA Contract DOT-FH-11-8738. January 1978.

Identifying and Analyzing Problems	2. Data Collection 3. Delay Computation 4. Environmental Impacts 5. User Simulation Models
Procedures for Designing and Evaluating Solutions	6. Detection of Recurring Congestion 7. Ramp Control 8. Mainline and Corridor Control 9. Bus/Carpool Priority Control 10. Detection of Non-Recurring Congestion (Incidents) 11. Incident Detection Algorithms 12. Planning for Incident Response 13. Cost-Effectiveness Evaluation Techniques 14. Simulation Models in Design and Evaluation
Implementing Solutions	15. Financial Planning and Staged Construction 16. Interagency Cooperation

SECTION GUIDE

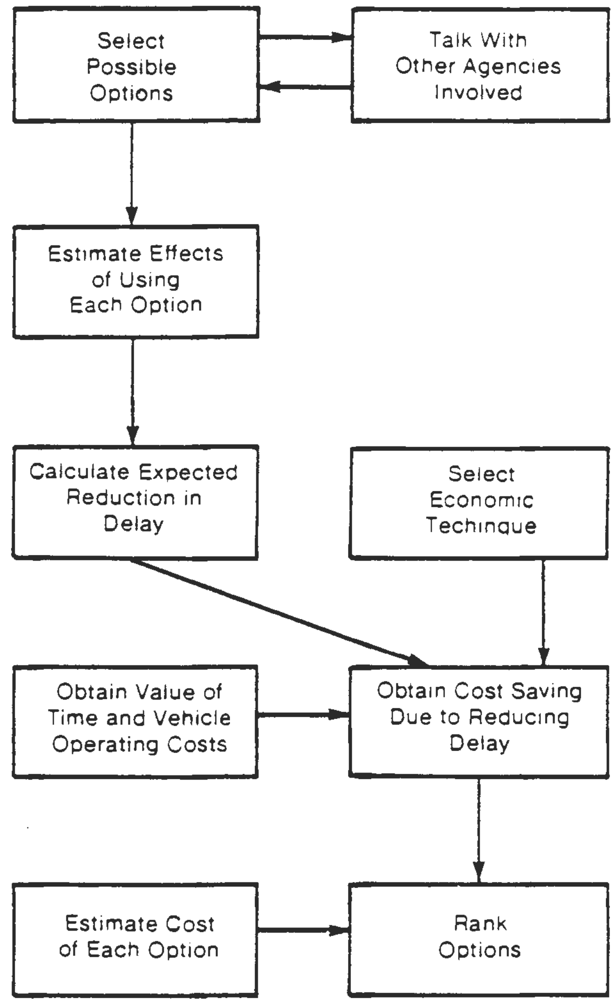
PROCEDURE FOR DESIGNING AND EVALUATING SOLUTIONS

The possible options for freeway management are discussed in chapters 6 through 12. Some of these chapters are dedicated to specific solutions such as ramp control (chapter 7), mainline and corridor control (chapter 8), of bus and carpool priority control (chapter 9). Other chapters indicate ways of identifying the location of recurring congestion (chapter 6) and non-recurring congestion (chapters 10 and 11). A general list of solutions is contained in chapter 12.

The involvement of other agencies is critical to the success of any freeway management system. Some of the issues that concern the interaction between agencies were identified in volume 1, others are included in the discussions located in chapter 12.

Chapter 3 provides guidance on the steps of the entire evaluation process which is illustrated in the figure on this page. Chapter 13 also discusses the pros and cons of different cost effectiveness evalution techniques that can be used in the analysis.

Chapter 14 explains how the simulation techniques of chapter 5 can be used to estimate delay reduction.



Evaluation Process

INTRODUCTION

Recurring congestion is a dynamic quantity which, as might be expected, is dependent upon the variations in flow that occur over time. For this reason it is important for the planners and engineers concerned with freeway management to know how the demand for each section of a freeway facility varies.

Variations in traffic demand are important, since they are necessary for:

- Calculating the extent of congestion due to a particular incident, geometric bottleneck, or overcapacity
- Determining the time periods for specific restrictive measures such as operation of a bus-only lane or a carpool facility
- Planning control strategies; for example, the calculation of the operational period for pretimed ramp metering

As traffic demand increases and approaches freeway capacity, the gap between vehicles is reduced. It is this reduction in gap that is the instigator of congestion. An individual driver will follow the vehicle ahead, leaving a gap that corresponds to his reaction time plus some safety factor (this safety factor is not stopping distance, but some time less than stopping distance). This minimum spacing is in the range of 1.2 to 1.4 seconds and is independent of traffic speed.

These minimum spacings correspond to a flow range of 2,600 to 3,000 vehicles per lane per hour. The question then arises as to why such flow levels are rarely found on freeways. The reason is that driver behavior is reflected in a variability of minimum spacings. Flow readings on freeways can reach 2,600 for short time periods, but a more reasonable figure for computing purposes would be 1,800 to 2,000 vehicles per lane per hour. This corresponds to an average minimum spacing of 1.8 to 2 seconds. As traffic flow increases and the average gap approaches the minimum value, drivers begin to influence each other by slowing down to maintain their own minimum spacing. When this occurs, the minimum spacing of the following driver is unacceptably reduced; he, in turn, slows down and this effect produces a chain reaction causing speeds to be reduced as flow increases. The dynamics are further complicated by lane-changing and a mixture of vehicle types. The resulting variation of speed with flow is a function of the form shown in figure 6.1.

Once the dynamics that cause congestion begin to operate, additional flow will create a large reduction in speed, making traffic move in a stop-and-go manner (corresponding to level of service F).

TEMPORAL VARIATIONS

The historical data used to accurately pinpoint recurring congestion problems should be complete, covering hourly, daily, and monthly variations in traffic flow. Variations within the peak hour should also be identified.

Hourly variations follow a reasonably predictable pattern for weekdays similar to that illustrated in figure 6.2. This example indicates an a.m. peak hour starting at 7 a.m. and containing 7.8 percent of the 24-hour volume, and a p.m. peak

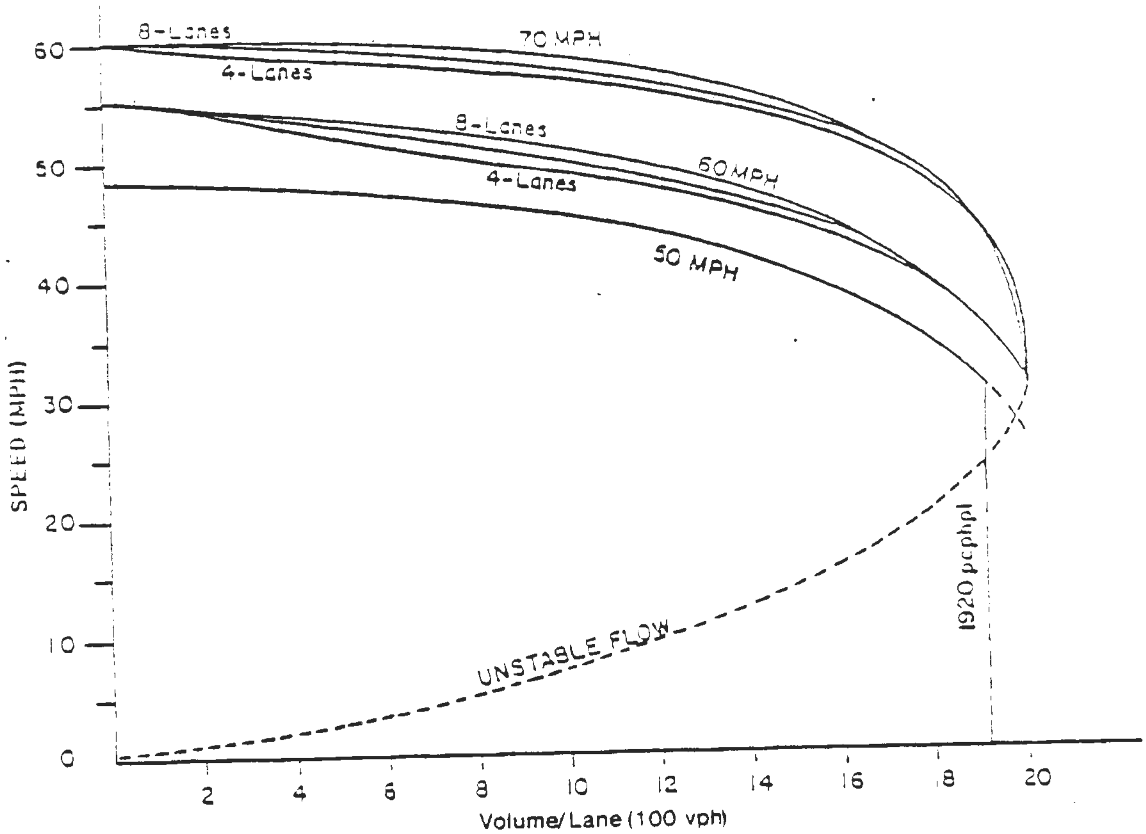


Figure 6.1. Speed-Flow Relationships Under Ideal Conditions

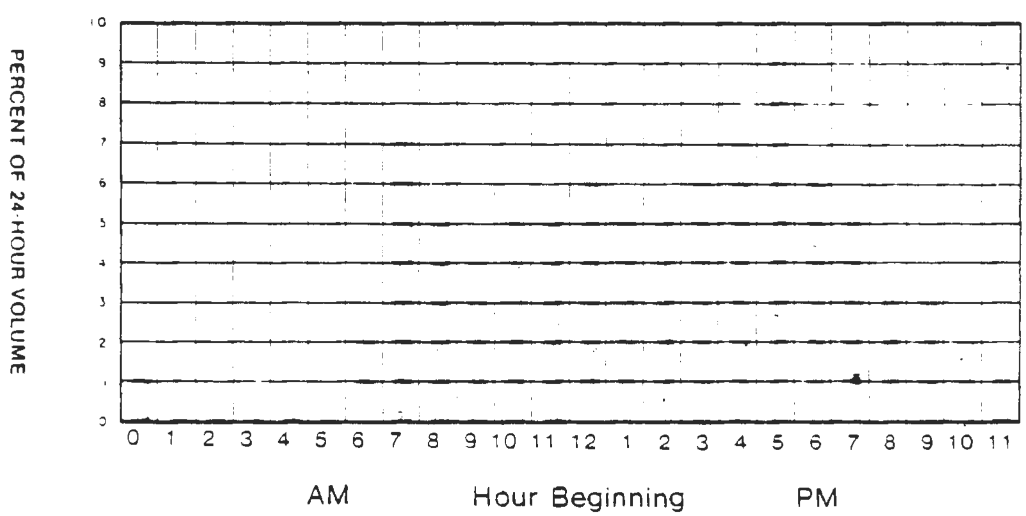


Figure 6.2. Hourly Variations of Traffic Flow

starting at 5 p.m. containing 9.5 percent of the 24-hour volume. There is also a minor peak starting at noon. This is the basic pattern of the hourly volume of traffic flow. Variations tend to occur according to geographical area, particularly in those cities with a central nucleus that has a large attraction for trip destinations. The principal variations on this pattern occur when:

- The time of the a.m. peak becomes earlier and the p.m. peak becomes later as the distance from the city center increases

- The dip between the a.m. and p.m. peaks is greater with increasing distance from the city center

- In areas close to the town center, this dip often effectively disappears and the facility has one long peak period from the morning to the evening

Variations within the peak hour can also be of importance to the freeway manager. The peak-hour factor for freeways is the ratio of the volume occurring during the peak hour to the maximum rate of flow during the 5- to 6-minute periods within the peak hour. If the freeway section under scrutiny is very near the point at which congestion will occur, a knowledge of the peak-hour factor can indicate to the freeway manager the extent of the likely congestion.

An example of this short-term peak and its effects on congestion are shown in figure 6.3. Although demand in this example exceeds capacity for 30 minutes, the resulting congestion lasts for 45 minutes. It is important to note that if only the hourly flow rates had been used for computing purposes, calculations would not have indicated any congestion.

Daily variations in traffic flow, like hourly variations, can be a useful guide to those planning a particular control strategy. Daily variations generally follow a pattern similar to that illustrated in figure 6.4. To a freeway manager, these data would indicate the extent to which various types of control should be modified on a day-to-day basis to anticipate the daily variations in congestion.

Monthly trends in traffic can also be used when contemplating control strategies. Figure 6.5 shows a typical pattern of monthly variations. These monthly variations should also be well formulated before any trend analysis is relied upon to predict traffic growth in future years.

GEOMETRIC CAUSES

Recurring congestion is often caused by the geometric design of a facility. Bottlenecks can result from lane reductions, horizontal or vertical alignment, ramp access, and excessive weaving. These are the common causes of recurring freeway congestion.

- Lane Reductions — Freeways often narrow when approaching urbanized areas, crossing bridges, or immediately after an off-ramp.

- Horizontal and Vertical Alignment — Horizontal curvature causes speed reduction that can quickly lead to congestion during heavy traffic flow. Vertical alignment has an effect on truck speed; this is particularly severe when the grade is steep.

- Free Ramp Access — Congestion can occur when an unrestrained ramp flow is allowed to access the main-

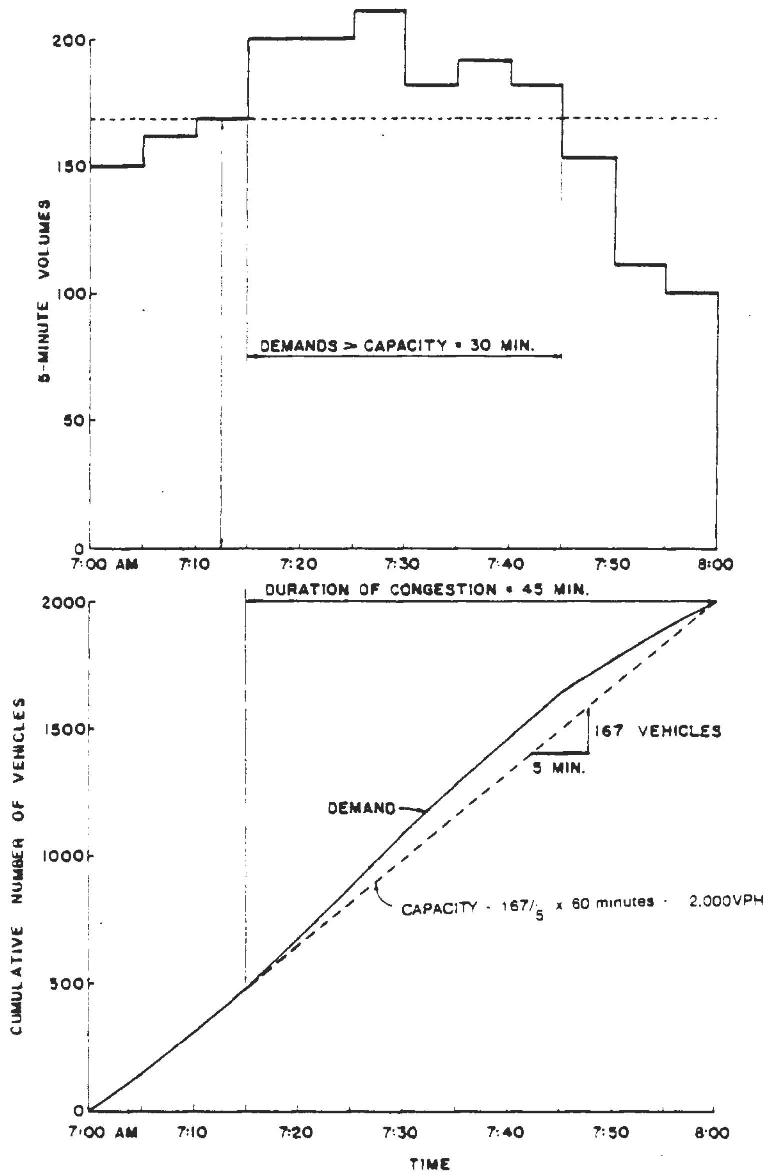


Figure 6.3. Relation Between Demand, Capacity, and Congestion (1)

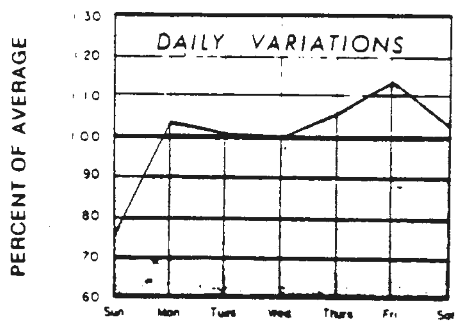


Figure 6.4. Daily Variations in Traffic Flow

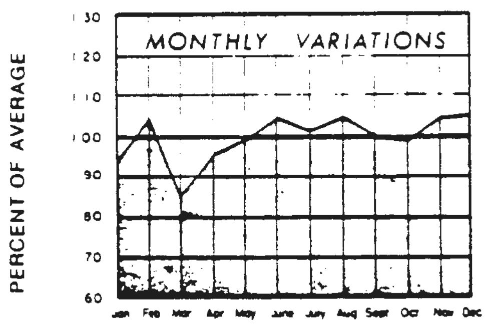


Figure 6.5. Monthly Variations in Traffic Flow

line flow. In peak periods the sum of these flows can be greater than the freeway capacity. The balancing of these flows to minimize delay is the basic principle of ramp metering, which is fully discussed in chapter 7 of this volume.

- Off-Ramps – Congestion from off-ramps occurs when an intersection downstream of the off-ramp has inadequate storage capacity. This causes the queue of exiting traffic to back up onto the freeway.

- Weaving – On- and off-ramps, located close together, can also give rise to congestion. In weaving areas, the capacity of the facility will be reduced because the speeds of weaving vehicles are lower than the speeds of through vehicles.

DETECTION METHODS

Determining the locations on which recurring congestion occurs on existing facilities is usually a straightforward process. Experience with the operation of the facility by the police, highway departments, and other agencies provides an important source of information in this effort.

Determining the volume levels at which congestion occurs and the severity of congestion are more complicated tasks, and can require sophisticated measurements of the characteristics of traffic flow on the freeway. These issues are discussed in the remainder of this chapter.

Vehicle Detectors

Detectors play an intensive role in freeway management systems. They provide the basic source of data that enables management systems to help prevent or reduce congestion. Figure 6.6 shows the possible locations of detectors at a freeway entrance ramp. The number of detectors shown in this figure is more than would normally be used at one ramp site, but this figure illustrates the most common uses of detectors and shows the geographical layout.

All detctors respond to the presence of a vehicle. It is the way in which this information is handled that distinguishes

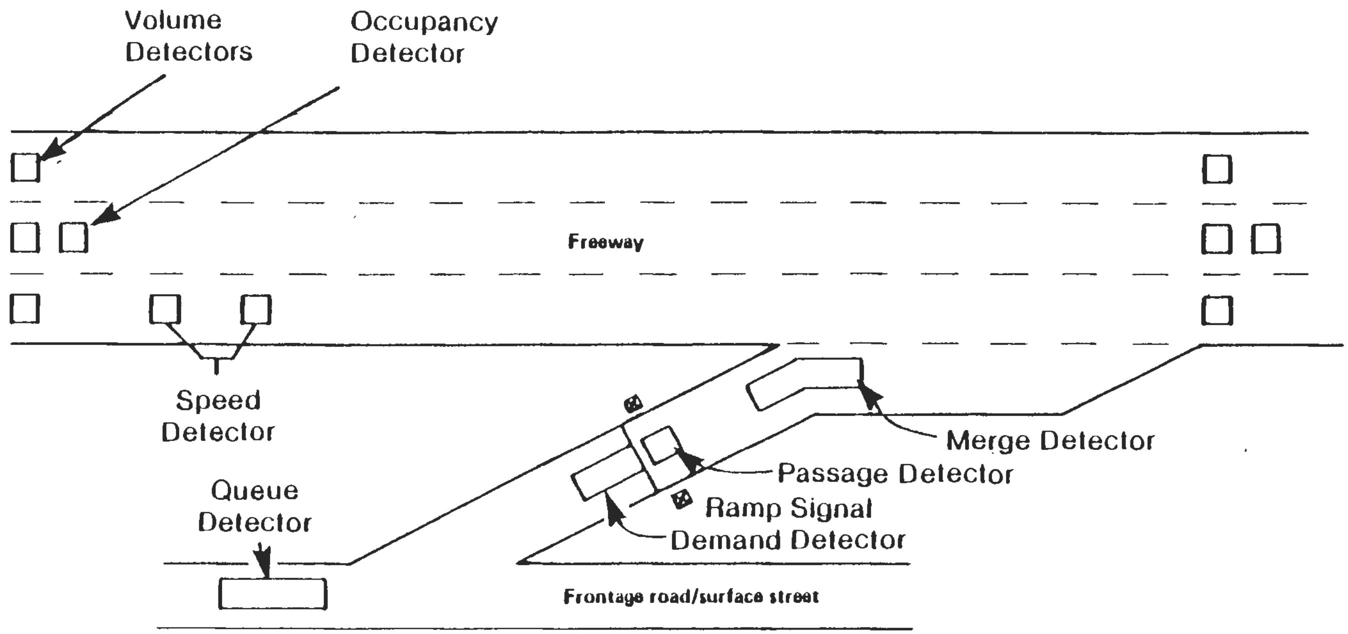


Figure 6.6. Possible Location of Detectors at a Freeway Entrance Ramp

between the different types of detector modes.

Volume detectors produce pulses that are counted, and the number of pulses correspond to the vehicle flow. Occupancy detectors provide a point measure of density; that is, the percent of time that the length of road covered by the detector is occupied. Presence detectors are used to determine if a vehicle is occupying the area.

These detectors can be used for both speed measurements, by using a pair of detectors, or queue measurements, by checking an individual location for occupancy. By combining detectors in a range of modes and geographic locations, traffic data can be collected and flow controlled. Details on detector types and their installation are discussed in chapter 12 of volume 3.

Other Data Sources

In addition to the automated systems previously described, there are often a variety of sources that provide information on freeway congestion. However, there is no substitute for going onsite and watching the mechanics of congestion being formed. Freeway engineers and planners should be actively involved in making personal observations. They may also consult a variety of other sources of information concerning the points where congestion regularly occurs, such as:

- The local traffic engineer
- Local radio and TV stations that use aerial surveillance to provide listener information
- Local citizen concerns and complaints
- Highway police who, by virtue of their often continual presence, are knowledgeable of local conditions

From sources such as these, an initial, but often non-quantifiable assessment of the locations and severity of regularly occurring congestion can be formed.

When congestion of a facility under construction is being investigated, existing parallel facilities can provide information on the various traffic patterns. However, a more comprehensive transportation planning or simulation exercise would be necessary to address the issues of recurring congestion.

REFERENCES

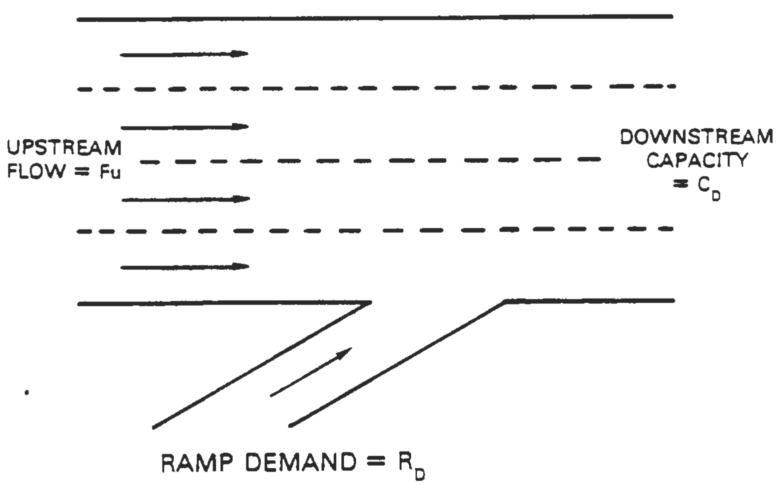
1. Drew, D. R., and Keese, C. J. "Freeway Level of Service as Influenced by Volume and Capacity Characteristics." Highway Research Record No. 99. 1965, pp 1-47.

INTRODUCTION

Ramp control is a method of balancing the upstream and downstream flows on either side of an entrance ramp. If the upstream flow plus the incoming flow on the ramp is greater than the downstream capacity, congestion will occur. By restricting the flow entering the ramp to the difference between the upstream demand and the downstream capacity, freeway congestion is reduced, and the total amount of delay to vehicles on the freeway and waiting to enter the freeway is reduced. This is the basis of ramp control, as illustrated in figure 7.1.

This simple theory leads to a series of considerations:

- What is the downstream capacity?
- Can the upstream demand be reduced?
- Will the queue caused by the waiting ramp traffic affect other traffic flow?



If: $Fu + R_D > C_D$

then congestion will occur.

If R_D is limited to a value of:

$C_D - Fu$

then congestion will be eased and maximum use made of the freeway.

Figure 7.1. Basic Ramp Metering

- Will any diverted ramp traffic cause the same problem by entering the freeway at the next available ramp?

- Since the upstream flow varies with time, should the ramp metering rate also vary with time?

- Is the metering rate so low that violations will occur?

These and other questions are addressed in the following sections that describe the various types of ramp control.

RAMP CLOSURE

Ramp closure consists of physically closing the ramp to all traffic. This may be a permanent closure or a short-term closure during peak periods.

Although ramp closure does not eliminate demand, it redistributes entering traffic to locations on the freeway where capacity is available to accommodate this demand. The freeway engineer considering ramp closure must be aware that this redistribution of demand will shift the problem to a different location.

Permanent Closure

Permanent closures may be necessary on older freeway systems where the original design did not adequately allow for the actual growth of traffic in specific areas. Many ramps on older freeway systems, particularly in downtown areas, were constructed with a lower design speed which involves a shorter merging distance. This often leads to reduced capacity at the merging location and produces congested conditions.

The methods that have been used to permanently close a ramp consist of either using signs or manually placing restrictive barriers across the ramp. Signs were found to result in a low compliance rate, and it is recommended that barriers be used to close a ramp permanently.

Temporary Ramp Closure

Temporary ramp closures may be a useful management device under any of the following circumstances:

- An incident has occurred on a ramp or immediately in the area of a ramp location, and the queue formed by vehicles trying to use the ramp is contributing greatly to congestion. (This technique can be used on both entrance and exit ramps.)

- Adequate storage is not available at an entrance ramp and the waiting vehicles are causing such a queue that the traffic on roads adjacent to the freeway is being substantially delayed.

- Although the inflowing ramp traffic should, ideally, equal the difference between downstream capacity and upstream demand, if the ramp metering rate is so restrictive that drivers will not comply with the ramp metering signals, then closing the ramp should be considered.

The alternative to ramp closure that has been used (in Toronto) is to stop the metering until the queue is reduced and then start it up again. Compliance levels by drivers at ramp signals are discussed later in this chapter.

The methods for temporary closure of ramps consist of automatic barriers for use when the time of closure (peak periods) is well defined. In the case of responding to incidents, temporary barriers would have to be placed manually. There are two types of automatic barriers that have been used successfully:

- The pop-up variety consisting of brightly colored cylinders set into the pavement, which rise from the ground when activated

- The conventional barrier, commonly found at toll booths

RAMP METERING PRINCIPLES

Metering is a method of restricting traffic flow(1). When applied as a form of entrance ramp control, metering is used to limit the rate at which traffic can enter a freeway. Metering rates used at an entrance ramp will usually be less than a practical maximum of 900 vehicles per hour (vph) and greater than a practical minimum of 180 to 240 vph.

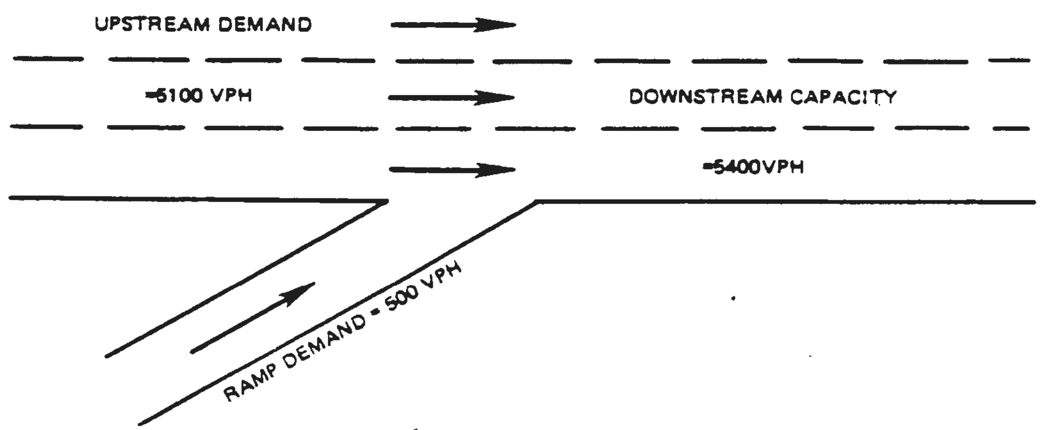
Metering Rates

The calculation of metering rates depends on the purpose for which the metering is being used. Normally, metering is used either to eliminate congestion on the freeway or to improve the safety of the merging operation.

Congestion — If the metering system is intended to eliminate congestion, the demand must consistently be less than capacity. Therefore, the calculation of the metering rate at a ramp would be based on the relationship between upstream demand, downstream capacity, and the volume of traffic desiring to enter the freeway at the ramp. Downstream capacity may be determined by the merging capacity at the ramp or by the capacity of the freeway section downstream.

Of course, if the sum of upstream demand and ramp demand were less than or equal to downstream capacity, metering would not be needed to prevent congestion. On the other hand, if the upstream demand alone is greater than downstream capacity, metering at the ramp would not eliminate congestion. Otherwise, the metering rate is set equal to the difference between upstream demand and downstream capacity.

For example, in the situation shown in figure 7.2, the upstream demand is 5,100 vph, the downstream capacity is 5,400 vph, and the ramp demand is 500 vph.



Ramp Metering Rate = Downstream Capcity - Upstream Demand
= 5,400 vph - 5,100 vph
= 300 vph
= 5 vpm

Figure 7.2. Example of Pretimed Entrance Ramp Metering Rate Calculation

Since the total demand (5,600 vph) is greater than the downstream capacity, congestion will occur on the freeway unless the total demand is limited to capacity. And, since the upstream demand is less than downstream capacity, ramp metering might be a feasible solution. Therefore, if a metering rate equal to the difference between upstream demand and downstream capacity (300 vph) is used, the freeway would be able to accommodate the upstream demand and maintain noncongested flow while also handling 300 vph or 5 vehicles per minute (vpm), of the ramp demand.

Manual calculations of ramp metering rates should be done in 15-minute time slices to account for variations in flow by time of day.

However, the ultimate test of the feasibility of ramp metering at a rate of 300 vph would involve the following considerations:

- Is adequate additional capacity available in the corridor for the 200 vehicles per hour that are likely to be diverted? And if so, is it likely that the 200 vehicles per hour would utilize that extra corridor capacity?

  If this is not likely, capacity would have to be added to the corridor and/or made more attractive for the 200 vehicles per hour that should be diverted. Otherwise, ramp metering would solve the problem on the freeway but worsen corridor congestion.

- Is adequate storage available at the ramp to accommodate the queue of vehicles that would have to wait at the ramp before entering the freeway?

  If it is not possible to provide adequate storage at the ramp, alternatives to be considered would be closure of the ramp, or metering at other ramps upstream to reduce upstream demand, which would, in turn, permit a higher metering rate and require less storage at the ramp.

- Is the specified metering rate (300 vph) too restrictive?

  If so, consideration should be given to closing the ramp or metering other ramps upstream to reduce upstream demand, which would permit a higher metering rate at the ramp. However, metering other ramps upstream might lead to underutilization of the freeway.

<u>Safety</u> — If ramp metering is to be used only as a means of improving the safety of the merging operation, then the metering rate is simply set at a maximum consistent with merging conditions at a particular ramp. The primary safety problem of the merging operation is incidence of rear-end and lane-change collisions caused by platoons of vehicles on the ramp competing for gaps in the freeway traffic system. Therefore, metering is used to break up these platoons and to enforce single-vehicle entry. To do this, the metering rate selected must be such that each vehicle has time to merge before the following vehicle approaches the merge area. The time it takes a vehicle to merge depends on the following factors:

- Distance the vehicle is stopped from the freeway

- Geometrics of the ramp (grade, sight distance, and length of the acceleration lane)

- Type of vehicle

- Availability of acceptable gaps in the freeway traffic stream

If the average time to merge is 6 seconds, the metering rate would be 10 vpm or 600 vph.

System Components

Pretimed metering is the simplest form of entrance ramp metering. Essential components of a pretimed metering system are:

- Ramp metering signal — Usually a standard three-color (red-yellow-green), or two-color (red-green), signal display that controls the ramp traffic
- Time clock actuated controlled — Frequently a standard one-dial or three-dial controller that regulates the timing of the ramp metering signal
- Advance ramp control warning signal — A sign that indicates to approaching traffic that the ramp is being metered

Depending on the purpose of the control and the strategy employed, the following devices are sometimes included:

- Check-in (demand) detector — Senses the presence of a vehicle at the ramp metering signal
- Check-out (passage)detector — Senses a vehicle leaving the ramp metering signal
- Queue detector — Senses the presence of vehicles waiting in a queue at the ramp metering signal

The location of these components on a ramp is illustrated in figure 7.3.

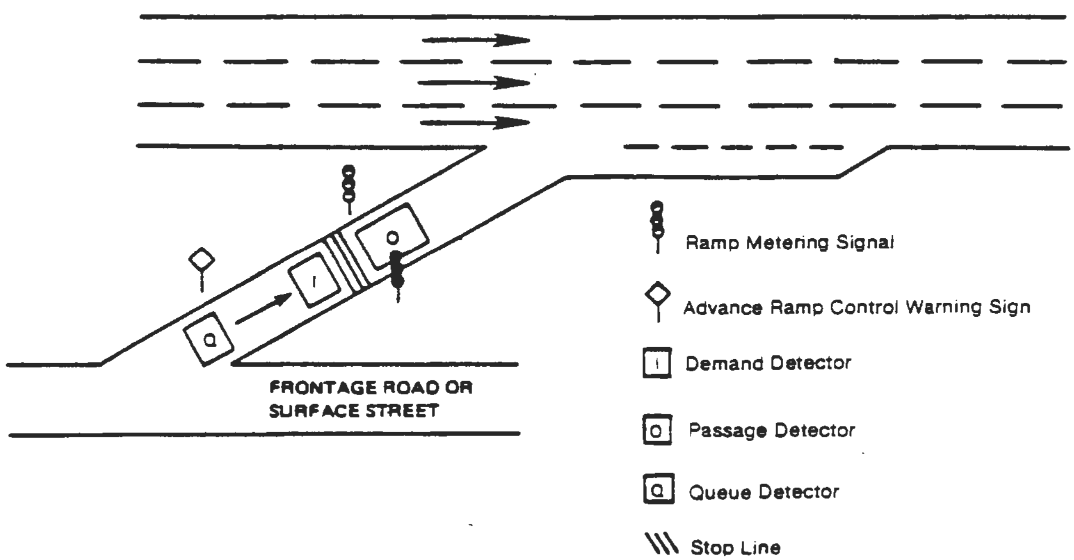


Figure 7.3. Layout of Pretimed Entrance Ramp Metering System

ITE GUIDELINES FOR RAMP METERING

To give the freeway planner or engineer guidance in the layout of ramps, the following section contains the Institute of Transportation Engineers recommended practice(2) for entrance ramp displays. The recommended practice stresses the need for uniformity: "Since the time has come that traffic control devices at freeway entrance ramps have been accepted as an operational tool to improve freeway operation, these differences must be resolved so that a true national standard can be adopted before too many installations of widely varied design have been built."

The recommended practices should, wherever possible, be used by engineers in order that over time a uniformity will be perceived by motorists. This standardization should lead to greater acceptance of such systems. These guidelines are currently under review and the reader should obtain a copy of the latest ITE recommendations.

Signal Display

1. Two signal heads, one on each of all metered ramps, are recommended. This configuration will provide more effective target value and a backup indication if one signal becomes inoperative or if one of the pedestals is damaged. Additional high-level or mastarm-mounted signal heads are permissible where vision of the signal is impaired by approach grades, curvature or queued traffic, or where other factors which require a high-level signal for improved visibility are present.

2. Recognizing that operational experience to date has been limited, it is recommended that the use of both two-color (red and green) and three-color (red, yellow and green) signals be permitted pending further investigation and research. Operational ramp metering signal installations to date have employed both two-color and three-color signal sequence for one-at-a-time metering operation. In general, the two-color signal systems have used demand and passage detector terminating the green indication. With this type of operation, the lengths of the green and red intervals are variable within a specified metering rate. The actuated operation compensates for differences in vehicle acceleration characteristics and driver reaction time, resulting in a more definite and responsive operation at the signal compared to a fixed-time signal sequence for one-at-a-time operation. The three-color signal systems have been operated, both actuated and fixed time, to establish the timing of the green and yellow intervals. In a fixed time system, the green and yellow timing is established for average conditions of vehicle acceleration and driver reaction. The use of the yellow interval, however, does provide additional flexibility in metering operation since the one-at-a-time and bulk metering modes may be used at the same location depending upon the control philosophy applied. In cases where only the bulk metering would be used, such as on some ramps whose acceleration lane becomes an additional lane of the freeway, a three-color indication must be used.

3. The use of 8-inch lenses mounted vertically is recommended as a minimum based upon unanimous satisfactory use. The use of larger lenses is permissible under exceptional circumstances but their use should be held to a minimum. Because of the lens position, it may be necessary in some installations to modify signal intensity through voltage reduction or some other suitable means.

4. Pedestal mounting of the signal head is preferred. Bottom height of the lowest lenses should be a minimum of 4.5 feet and a maximum of 6 feet above the roadway. This mounting configuration is required to draw the approaching vehicles into a closely controlled area of influence. Attention should be given to truck driver eye height; mounting or aiming adjustments may have to be made.

5.a. Ramp metering signal should not be blank during period of nonmetering operation. The signal should be live at all times to give the motorist a positive indication that the signals are functioning. Steady green is the concensus candidate for dwelling or nonmetering operation, but we feel that it has potential shortcomings:

- Traffic becomes accustomed to moving past the signal and is surprised by the red.
- The first red indication of the metering period presents a potential for rear-end collisions.
- Accident experience during the off-peak periods has not been adequately evaluated.

In spite of the possible shortcomings, the use of the steady green indication is recommended with the stipulation that the rear-end accident potential must be recognized when the first red indication is displayed.

5.b. The transition from dwell or nonmetering operation to the first red indication should be:

- In a system where the yellow indication is not used, control logic with advance detection to ensure that no vehicle is in close proximity to the signal should be employed.
- Where a three-color signal display is used, a steady yellow indication should precede the red. Flashing beacons, used in conjuction with advance warning signs and actuated shortly before the first red indication, may also be used.

6.a. Signal location from the ramp nose should be based upon vehicle acceleration characteristics from a stop at the signal to attain a safe merging speed at the end of the usable portion of the acceleration lane. This must be given special attention in exceptional cases such as wide ramps at which two-abreast metering is permitted. Consideration must also be given to adequate storage distance between the signal and the service drive or crossroad. It can be seen from these considerations that not all ramps are geometrically capable of being metered. In some special cases it may be feasible to use a nearby intersection signal to meter ramp traffic. While the foregoing comments have been intended as a guideline for the implementation of ramp metering of existing facilities, consideration for future ramp metering should be provided for in the design of new facilities and the reconstruction of existing facilities.

6.b. Lateral placements of the signals should be as close to the roadway as practicable with lateral clearance adequate to protect the device from accidental damage by traffic. Since the operational requirements of these signals require them to be close to the roadway, it is recommended that breakaway features be incorporated into the design of the pedestals.

6c. Signal heads should be aligned in such a manner that they can be seen from the stopped position at the point of obedience. The right-hand signal head should be aimed up the ramp to favor approaching traffic with maximum sight distance, while the left-hand signal should favor the driver while he is stopped at the point of obedience.

7. Portable signals are not recommended and should not be used. For experimental systems, consideration should be given to use of temporary, but not portable, devices.

8. Flashing yellow beacons used in conjuction with advance warning signs should be permitted but not required. Most present applications were new research and possibly installed because of excessive caution.

9. The option of using pretimed versus traffic-responsive signals should be local and based on the availability of funds and the degree of sophistication desired or required. In some instances, adequate benefits for today's traffic may be derived through the use of less sophisticated forms of metering, such as with a local controller handling a single ramp in a preset manner. Such an installation of a computer-controlled network may not prove cost effective in solving today's problems. If such a system of inexpensive ramp metering is proposed as the immediate solution to a problem, the hardware should be compatible with a computer-controlled surveillance and control network since it is likely that additional levels of sophistication may be desired later.

Signs

1. Research on the use of changeable message advance information signs, which are intended to advise motorists of freeway conditions or of preferred alternative routes to the freeway or through the corridor, is not complete with respect to types, application, and effectiveness. No firm recommendation is offered at this time.

2. Advance warning signs, those intended to alert traffic to the presence of a signal on the ramp and to the fact that they may be required to stop on the ramp, should be used near the head of the ramp whenever metering signals are present. They should be placed in a position that affords good target value for traffic entering the ramp that allows for adequate evaluation and reaction time, yet do not obscure the metering signals from the view of approaching traffic. In some instances, placement on the left side of the ramp may be desirable. If metering is to operate during hours of darkness, the signs should either be reflectorized or illuminated.

3. Signs to supplement a stop bar or to delineate the point of obedience may be desirable in certain cases. When used, the legend "wait here for green" with an arrow pointing toward the stop bar or point of obedience is recommended. The sign should conform to standard regulatory sign specifications. Supplementary signs advising "one car only on green" and similar legends appear to be incompatible with the recommended continuous green "not metered" condition, and preliminary investigations seem to indicate that such signs may not be required with two-color signals using a short green time. Further study is needed to determine whether signing of this sort is effective or necessary. One objective of the suggested further study would be to examine the need of such signs for enforcement purposes.

4. The use of ramp metering should not preclude the use of "merging traffic" signs on the freeways where they would normally be used. These signs are considered a part of main freeway signing and relate to the caution that must be excercised both by the ramp traffic and by the freeway traffic during the act of merging.

5. Other signs bearing supplemental information such as the hours during which metering would normally be operated may be desired locally. These forms of signing are not recommended, but are not ruled out or prohibited. They should be used sparingly and judiciously, however, since the driver's attention must be divided between his driving task and all of the control devices; an excessive number of these devices detracts from the effectiveness of each of them.

Pavement Marking and Channelization

1. It is recommended that a stop bar of 12 to 24 inches width be placed between zero and no more than 6 feet in advance of the signal. Because of the close time tolerances under which one-at-a-time metering operates, it is necessary that vehicles be brought into a closely controlled area of influence. When it is anticipated that metering will be operated during hours of dusk or darkness, the stop bars should be reflectorized.

2. The effectiveness of other pavement markings and channelization in the forms used to date has not been conclusively demonstrated. Each location should be evaluated on its individual merits; and markings or devices that are locally effective and consistent with good engineering practice, should be employed as necessary.

Additional Considerations

Several items relating to entrance ramp displays deserve mention, although detailed discussion is not within the scope of this report. Of particular interest are the following:

1. It is recommended that individual state vehicle codes be amended specifically to include ramp metering controls. Because the ramp metering situation deviates from the intersectional traffic control application, it is felt that an adequate legal basis should be established for ramp control use.

2. Amendments to the state manual of uniform traffic control devices to include ramp metering devices are also suggested. Ramp control type placement and operation should be uniform.

Detectors — Depending on the degree of equipment sophistication deemed essential to perform a particular control strategy, the detectors required for ramp metering can include:

1. A ramp signal demand detector located immediately in advance of the control displays to detect the vehicle stopped nearest the red signals. Presence in the demand detection zone is required to produce a red-to-green signal change. If used, the size and location of this detection zone is critical since failure to detect the lead vehicle in a queue will cause serious system malfunction.

2. A ramp signal passage detector located immediately beyond the control displays to detect vehicle passage through the ramp signals. Presence in the passage detection zone is required to produce a green-to-red

signal change. Clearance of the passage detection zone may be required before the control logic considers the next vehicle in the demand detection zone. Actuation of the passage detector when the ramp signals are red indicates a signal violation and enables compliance checks.

3. A ramp queue detector located to detect storage behind the control signals or spillback into an adjacent frontage road or intersection. Prolonged presence in the queue detection zone may cause the metering rate to allow more vehicles through the ramp signals, if the control strategy so provides. In system strategies, any increase in the metering rate at a saturated ramp may be offset by corresponding decreases at adjacent interchanges.

   This detector also can be used to help safely change the metering device from the nonmetering mode to the initial rest-in-red condition at the beginning of the metering period when no traffic is detected approaching the ramp signals.

4. A ramp merge detector located to detect storage in the merging area of the ramp, particularly where poor merge geometrics exist. Prolonged presence in the merge detection zone may cause additional ramp vehicles to be held at the metering signals, if the control strategy so provides.

   Due to the precise requirements for detection zones to cover all ramp vehicles, most existing installations use an induction loop presence detector. The detection zone can be designed to cover any portion of the ramp width usable by vehicles. Channelization techniques may be necessary where one-lane operations are desirable, particularly on wide ramps.

   Although up to four detectors would be the usual ramp detection requirement, one or more additional detectors are usually needed on the freeway through-lanes and on exit ramps to provide inputs to whatever strategy is desired for changing ramp metering rates based upon freeway traffic flow conditions.

<u>Control Strategies</u> — There are numerous strategies which have been developed for controlling the ramp signal change timing, ranging from simple local pretimed methods which attempt to keep ramp volumes lower than expected available freeway capacity, to complex computer-controlled methods which handle several ramps in traffic-responsive schemes featuring multiple control programs and over-rides. Generally, the major benefits from ramp controls are realized by providing some form of ramp control in lieu of no control. As refinements are added to increase the sophistication of the ramp control strategy, the detection equipment requirements tend to increase while the incremental benefits tend to decrease, thereby suggesting careful system cost-effectiveness design.

The most common control strategies in use today are based on local systemwide application of the following three types, or variations thereof:

1. Pretimed metering: Rates are based on historical data relating ramp flows to freeway demands and capacity.

2. Traffic-responsive metering: Rates are based on real-time data, measuring estimates of available freeway capacity and/or other traffic flow measurements such as lane occupancy.

3. Traffic-responsive merge control: Individual vehicles are released to meet right-lane gaps measured up-

stream and projected to be available and acceptable at the merge area. Some sophisticated systems are advanced algorithms with the moving merge concepts.

Whatever the control strategy, the control function at each ramp may be performed by a local controller, possibly interconnected with other ramps, or by a central digital computer, which may handle system surveillance functions as well as system control functions. The selection of an appropriate control strategy should consider the freeway and ramp geometrics and traffic conditions to determine the cost-effectiveness of the ramp control portion of various equipment configurations for accomplishing particular strategies.

Other Considerations — Other considerations that may affect the nature of entrance ramp displays include: special geometric and visibility problems; two-lane ramp operation; priority bus ramp use; channelization techniques; and local maintenance and traffic requirements.

Lighting and Reflectivity — Because of the critical nature of visibility as it relates to proper ramp operation, consideration should be given to lighting of signs and reflectorization of signs and critical objects and markings along the ramp. Use of higher index beads for paint reflectorization may at times be desirable.

PRETIMED OPERATION

In the operation of a pretimed metering system, the ramp signal operates with a constant cycle in accordance with a metering rate prescribed for the particular control period. However, timing the red, yellow, and green intervals of the cycle (some systems use ramp signals that have only red and green intervals) depends on the type of metering used: single-entry metering or platoon metering.

Single-Entry Metering

In the case of single-entry metering, the ramp metering signal is timed to permit only one vehicle to enter the freeway per green interval. Therefore, the green-plus-yellow (or just green if yellow is not used) interval is just long enough (usually about three seconds) to allow one vehicle to proceed past the signal, and then the remainder of the metering cycle is red. For instance, if a metering rate of 600 vph or 10 vpm was used, the green-plus-yellow interval would be 3 seconds and the red interval would be 3 seconds. If the metering rate was 300 vph, or 5 vpm, a green-plus-yellow interval of 3 seconds and a red interval of 9 seconds would be used.

Platoon Metering

When metering rates greater than 900 vph are required, platoon metering, which permits the release of two or more vehicles per cycle, must be used to achieve such high metering rates. For pretimed platoon metering, the cycle length is determined on the basis of the desired metering rate and the average number of vehicles to be released per cycle. For example, in the case of a metering rate of 1,080 vph, or 18 vpm, and a release of two vehicles per cycle, nine cycles per minute would be required. Therefore, the cycle length would be 6.67 seconds. Similarly, if a release of three vehicles per cycle was used instead, the cycle length would be 10 seconds. However, the timing of the cycle intervals (i.e., green, yellow, red) depends on the form of platoon metering used: tandem or two-abreast.

Tandem Metering — In the case of tandem metering, the vehicles are released one behind another. Therefore, the green-plus-yellow time is made long enough to permit the clearance of the desired number of vehicles per cycle. A yellow interval should be used to minimize the rear-end collision potential. Thus, for the 6.67-second cycle with two-vehicle platoons, a 4.67-second green-plus-yellow and a 2-second red might be used. And for the 10-second cycle with three-vehicle platoons, a 7-second green-plus-yellow and a 3-second red might be used. Experience indicates that two-vehicle platoons can be handled satisfactorily and that three-vehicle platoons are a practical maximum; and that in either case a maximum metering rate of 1,100 vehicles per hour can be expected.

<u>Two-Abreast Metering</u> – With two-abreast metering, two vehicles are released side by side per cycle. This form of metering requires two parallel lanes on the entrance ramp and a sufficient distance beyond the ramp metering signal for the two vehicles to achieve a tandem configuration before merging with freeway traffic. The timing of the cycle intervals for two-abreast metering is similar to that for single-entry metering in that the green-plus-yellow interval is just long enough (usually about 3 seconds) to allow one vehicle in each lane to proceed past the ramp metering signal. The remainder of the cycle is red. Maximum metering rates of about 1,100 vph may be achieved with two-abreast metering. Experiments in which the two vehicles have been released simultaneously, as well as alternately, have shown no appreciable advantages of one operation over the other.

Compared to single-entry metering, platoon metering has a number of disadvantages: greater driver confusion, greater probability of rear-end accidents, and greater possibility of disrupting freeway flow. Therefore, single-entry metering should always be given first consideration, and platoon metering should not be used unless it is absolutely necessary to achieve higher metering rates. Of the two forms of platoon metering, two-abreast metering is generally preferred, because it usually causes less driver confusion and provides a safer operation.

- <u>Check-in (demand) detector</u> – In some applications of pretimed metering, a check-in detector is placed on the approach to the ramp metering signal so that the signal will remain red until a vehicle is detected at the stop line (refer to figure 7.3). When a vehicle is detected by the check-in detector, the ramp metering signal will change to green provided the minimum red time has elapsed. With this type of operation, it is desirable to set the signal for a minimum metering rate (e.g., 3 vpm) in case there is no detector actuation (due to detector failure or because of vehicles stopping too far back from the stop line to actuate the detector).

- <u>Check-out (passage) detector</u> – In some systems, a check-out detector has been used to ensure single-vehicle entry. When a vehicle is permitted to pass the ramp metering signal, it is detected by the check-out detector. The check-out detector is installed just beyond the stop line (usually about 8 feet past it). The green interval is terminated as soon as the vehicle is sensed by the check-out detector. In this way, the length of the green interval is made sufficient for the passage of only one vehicle.

- <u>Queue detector</u> – In some pretimed metering systems, a queue detector is used to prevent ramp traffic from blocking frontage roads or surface streets. The queue detector is placed at a strategic point on the ramp, or on the frontage road, in advance of the ramp metering signal. When it senses a vehicle, which indicates that a queue of vehicles waiting at the ramp metering signal has reached a point where it is likely to interfere with traffic on the frontage road or surface streets, a higher metering rate is used to shorten the queue length.

Benefits

Benefits of pretimed metering systems, when compared to the situation without control, are usually expressed in terms of the following:

- Shorter traveltimes
- Less delay
- Higher freeway speeds
- Higher freeway service volumes
- Safer merging operation (fewer rear-end and lane-change collisions)

Benefits that have been reported(3) for some pretimed metering installations are summarized in table 7.1. It has been shown that this type of entrance ramp control can be extremely cost-effective. Experience in California has indicated that the biggest net gain in benefits is realized in going from no control to pretimed metering, and that any increases in control sophistication quickly run into diminishing returns.

TRAFFIC-RESPONSIVE OPERATION

Although pretimed metering gives the driver a dependable situation to which he can readily adjust, a major disadvantage is that the system cannot respond to significant changes in demand or adjust to unusual traffic conditions resulting from incidents. Because of this inability to respond to changes in traffic conditions and the relative difficulty of dissipating congestion, pretimed metering rates have usually been set so that operation would

TABLE 7.1. SUMMARY OF REPORTED BENEFITS OF PRETIMED ENTRANCE RAMP METERING SYSTEMS

Location	Number of ramps used in evaluation	Traveltime: Change (Veh.-hrs/yr.)	Traveltime: Annual savings	Other effects	Assumptions
St. Paul, Minnesota	2	-5,640	+16,900	Corridor throughput unchanged. Slight reduction in alternative route travel times.	130 dry working days per year. No benefits on wet days. Value of time equal $3 per veh.-hr.
Los Angeles, California	6	-108,300	+325,000	Additional benefits through increased driver satisfaction, less frustration, etc.	130 incident-free p.m. peak periods per year. No benefits on days on which incidents occur. Value of time equal $3 per veh.-hr.

be at volumes slightly below capacity. To overcome the disadvantage of pretimed metering being unresponsive to upstream demand, traffic-responsive ramp control was developed. The basic strategy of traffic-responsive metering is as follows:

- Obtain real-time measurements of traffic variables on the freeway
- On the basis of these measurements, determine where the freeway is operating on the fundamental speed-flow curves
- Determine the maximum ramp volume of vehicles that can be allowed to enter the freeway without causing congestion

There is a series of basic forms of traffic-responsive metering depending on which function of the traffic flow characteristics is used as the criterion for control:

- Demand-capacity control
- Occupancy control
- Shoulder-lane space control
- Gap acceptance merge control
- Gap acceptance/demand-capacity control

The following sections describe how these various methods of traffic-responsive ramp metering work. For installation details, the reader is referred to volume 3. In situations where the evaluation, by simulation, of a series of metered ramps is required, this is described in chapter 5 of this volume.

Demand-Capacity Control

Demand-capacity control features the selection of metering rates on the basis of a real-time comparison of upstream volume and downstream capacity. The upstream volume is measured in real time and compared with either a preset value of downstream capacity determined from historical data or real-time value computed from downstream volume measurements. To be most effective, the downstream capacity used should account for the effects of weather conditions, traffic composition, and incidents on capacity.

The difference between the upstream volume and the downstream capacity, referred to as available capacity, is then determined and used as the allowable entrance ramp volume. The allowable entrance ramp volume is expressed as a metering rate to be used during the next control interval (usually 1 minute). If the upstream volume is greater than the downstream capacity, a minimum metering rate is used (e.g., 3 to 4 vpm). Of course, theoretically, if the upstream volume were greater than the downstream capacity, a zero metering rate, or ramp closure, should be used in order to prevent congestion. It has been found that, generally, metering rates lower than 3 vpm are not effective, because vehicles waiting at the ramp will judge the ramp metering signal to be malfunctioning and proceed through on red. Also, a positive form of ramp closure is not practical for the short control intervals that are used in traffic-responsive metering.

Since a low upstream volume could occur in congested or uncongested flow, volume alone does not indicate degrees of congestion. Therefore, an occupancy measurement is usually made to determine whether uncongested or congested flow prevails. If the occupancy measurement is above a preset value that is determined from historical data, congested flow would be assumed to exist and a minimum metering rate used.

Occupancy Control

Occupancy control utilizes real-time occupancy measurements taken upstream or downstream of the entrance ramp. One of a number of predetermined metering rates is selected for the next control interval (usually 1 minute) on the basis of occupancy measurements taken during the current control interval. For a given entrance ramp, the metering rate to be used for a particular value of occupancy would be based on a plot of historical volume-occupancy data collected at each measurement location.

An example of a typical plot is shown in figure 7.4. From such a plot, an approximate relationship between volume and occupancy is established, and a value of occupancy at capacity is determined. Thus, for each level of occupancy measured, a metering rate that corresponds to the difference between the predetermined estimate of capacity and the real-time estimate of volume can be computed. If the measured occupancy is greater than or equal to the preset capacity occupancy, a minimum metering rate would be selected instead of a zero rate or ramp closure. This choice would be based on practical entrance ramp control considerations.

Shoulder Lane Space Control

This technique is a variation on demand capacity control in which only the shoulder lane is used in calculating the upstream demand. This demand is then compared with the capacity of the shoulder lane that is downstream of the intersection and the ramp metering rate calculated accordingly. Although this technique is cheaper because detectors are only required in one lane, it has a distinct disadvantage in that it fails to allow for vehicles changing from the shoulder lane into the lane on the left. This is a common maneuver as drivers in the shoulder lane can see the ramp traffic entering the system and change lanes accordingly. Because of this, shoulder-lane control tends to overcount the upstream demand and this results in a lower than optimum ramp metering rate.

Gap Acceptance Merge Control

Figure 7.5 indicates a layout of detectors in the area of a freeway ramp. Gap acceptance merge control involves the incoming vehicle on the ramp fitting into an acceptable gap in the shoulder-lane traffic. The speed detectors (as shown in figure 7.5) are used to detect a gap above a preset acceptable level and then release a vehicle from the ramp metering signal such that it merges into the acceptable gap. The traveltime of the entering vehicle on the ramp has to be matched with the traveltime of the gap in the traffic. This technique is not in general use.

Gap Acceptance and Demand-Capacity Control

This technique combines both demand-capacity and gap acceptance techniques. The demand-capacity technique is used to establish a metering rate as explained earlier. However, within the limits of that metering rate, individual vehicles are released from the ramp to fit in with acceptable gaps in the traffic.

Operational Considerations

These are the basic techniques for traffic-responsive ramp metering; there are, however, certain operational considerations that apply to all the forms of control:

- Continuous occupancy of the presence detector in the merging area (indicating a particularly hesitant driver or an incident) can be used to hold vehicles at the ramp signal until the merging area is clear.

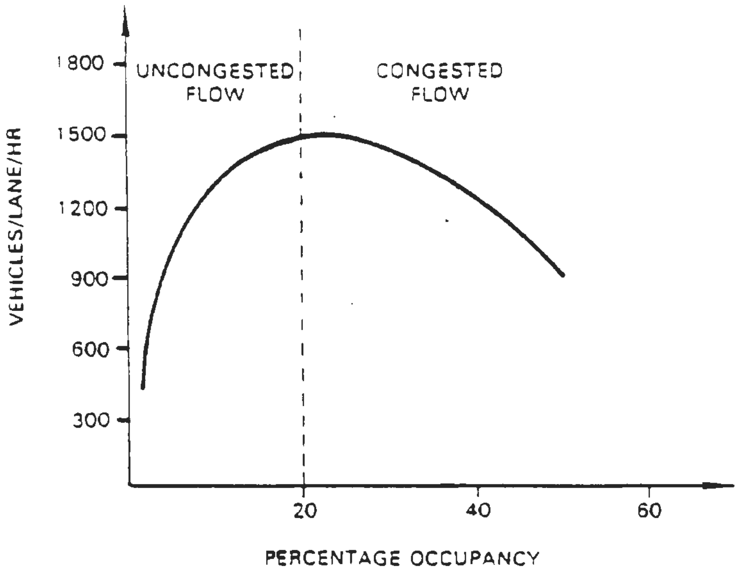


Figure 7.4. Typical Volume - Occupancy Plot

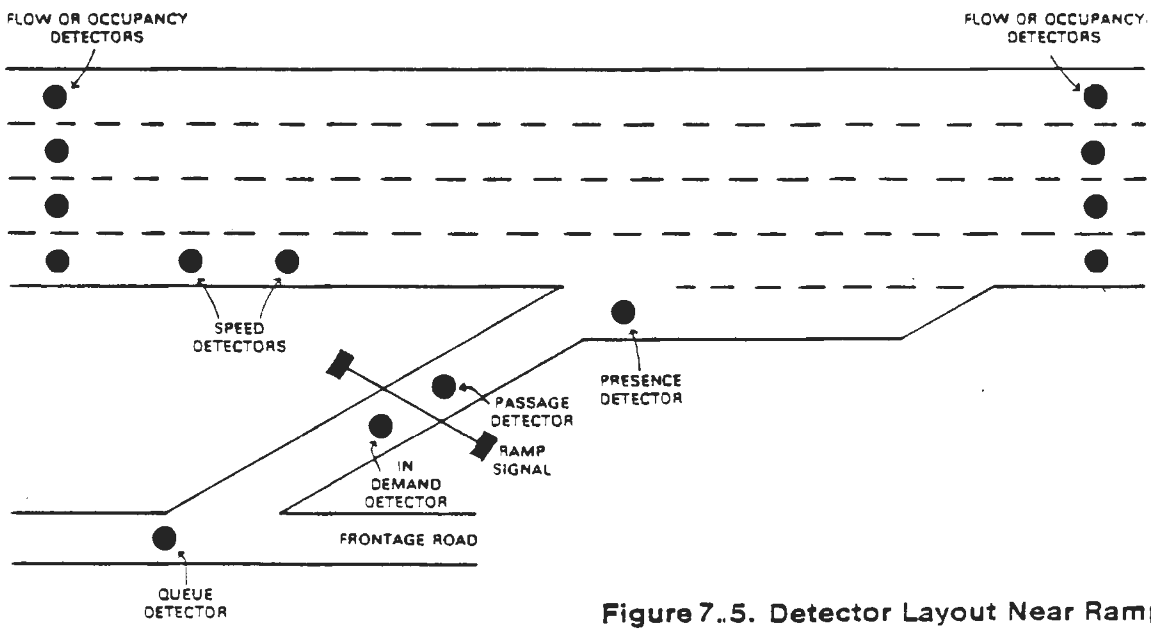


Figure 7.5. Detector Layout Near Ramp

- Continuous occupancy of the queue detector indicating that the line of vehicles waiting to use the ramp has reached back to the queue detector could be alleviated by increasing the ramp metering rate or giving the ramp traffic a long green. This decision would depend upon local conditions and how important a blockage on the frontage road is to local traffic.

- If non-operation of the passage detector indicates that a waiting vehicle has missed the green signal, a second signal should be given.

- If the queue detector is actuated but the demand detector is not actuated within a reasonable time, the vehicle could have stopped short of the demand detector. Providing a green signal should clear the vehicle from this position.

The engineer will have little difficulty in the calculation of an effective ramp metering rate for an individual ramp location. When more than one ramp is to be controlled, however, these calculations cannot be considered in isolation. Altering one ramp affects the other ramps in the system. The technique to account for this is termed integrated ramp control.

INTEGRATED RAMP CONTROL

Integrated ramp control refers to the application of ramp control to a series of entrance ramps where the interdependency of entrance ramp operations is taken into account. The primary objective of integrated ramp control is to prevent freeway congestion. Therefore, the control of each ramp in an integrated ramp control system is based on the available capacity considerations for the whole system rather than on the available capacity constraint at each individual ramp.

If congestion is to be prevented on the freeway system, the concept of integrated ramp control must be used in the design of a system of entrance ramp controls for any section of freeway with more than one entrance ramp. It is applied in the following types of systems:

- Pretimed metering (including ramp closure)

- Traffic-responsive metering

- Gap acceptance merge control

A discussion of integrated ramp control applied to each of these systems follows.

Integrated Pretimed Metering

Integrated pretimed metering refers to the application of pretimed metering to a series of entrance ramps. The metering rate for each of these ramps is determined in accordance with available capacity constraints at the other ramps as well as its own local available capacity constraint. These metering rates, which are computed from historical data pertaining to each control interval, require the following information:

- Mainline and entrance ramp demands

- Freeway capacities immediately downstream of each entrance ramp

- Description of the traffic pattern within the freeway section to be controlled

This information provides the basis for establishing the available capacity constraints of the entrance ramps and their interdependencies.

Fundamental Metering Rate Calculations

Given the required data, the fundamental procedure for computing metering rates involves five steps:

Step 1: Start with the entrance ramp which is farthest upstream.

Step 2: Determine the total demand (upstream mainline demand plus ramp demand) for the freeway section immediately downstream of the ramp.

Step 3: Compare the total demand to the capacity of the downstream section, and proceed as follows:

- If the total demand is less than the capacity, metering is not required at this ramp by this available capacity constraint. Therefore, skip step 4 and go immediately to step 5.

- If the total demand is greater than the capacity, metering is required at this ramp by this available capacity constraint. Therefore, proceed to step 4.

Step 4: Compare the upstream mainline demand to the capacity of the downstream section, and proceed as follows:

- If the upstream mainline demand is less than the capacity, then the allowable entrance ramp volume (or metering rate) is set equal to the difference between the capacity and the upstream mainline demand.

- If the upstream mainline demand is greater than or equal to the capacity, then the allowable entrance ramp volume is zero, and the ramp must be closed. If the upstream mainline demand is greater than the capacity, the volumes permitted to enter at ramps upstream must be reduced accordingly. The total reduction in the allowable entrance ramp volumes upstream is equal to the difference between the upstream mainline demand and capacity, adjusted to account for that portion of the traffic entering upstream that exits before it reaches the downstream entrance ramp being closed.

Step 5: Select the next entrance ramp downstream, and go back to step 2.

This procedure is illustrated by the following examples.

Example Number 1 — In the example shown in figure 7.6, pretimed metering rates are calculated for an integrated, pretimed control system comprised of four entrance ramps. In reviewing this example, the following points should be noted:

Since only entrance ramp control is being considered and not mainline control, the allowable mainline volume X is set equal to the mainline demand D.

Using the notation given in figure 7.6, the demand S_j at a section j is computed by the following equation:

$$S_j = \left[\sum_{i=1}^{j} A_{ij} X_i \right] + A_{j+1,j} D_{j+1}$$

Where:

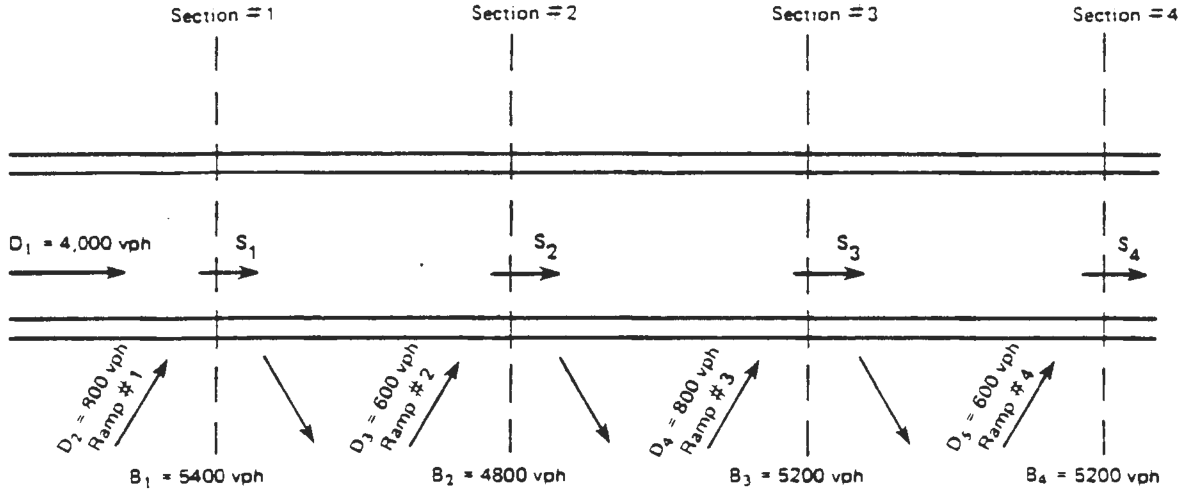
X_i = allowable volume at input_i

D_i = demand at input_i

A_{ij} = decimal fraction of vehicles at input i which pass through section j

S_j = demand at section j

As it happens, the metering rate computed for each entrance ramp in this particular example is determined solely by the available capacity constraint at the section immediately downstream and is not influenced by the available capacity constraints at other ramps.



X_i = allowable volume at input i

D_i = demand at input i

B_j = capacity at section j

A_{ij} = decimal fraction of vehicles entering at input i which pass through section j

S_j = demand at section j

A_{ij} Values

i \ j	1	2	3	4
1	1.00	0.95	0.90	0.85
2	1.00	0.75	0.70	0.60
3	–	1.00	0.90	0.85
4	–	–	1.00	0.90
5	–	–	–	1.00

<u>Compute X_i's starting at section #1</u>

- Set $X_1 = D_1 = \underline{4000 \text{ vph}}$
- $S_1 = A_{11}X_1 + A_{21}D_2 = (1.00)(4000)+(1.00)(800) = 4800 \text{ vph} < B_1 = 5400 \text{ vph}; \therefore \underline{X_2 = 800 \text{ vph}}$
- $S_2 = A_{12}X_1 + A_{22}X_2 + A_{32}D_3 = (0.95)(4000)+(0.75)(800)+(1.00)(600) = 5000 \text{ vph} > B_2 = 4800 \text{ vph}; \therefore \underline{X_3 = 400 \text{ vph}}$
- $S_3 = A_{13}X_1 + A_{23}X_2 + A_{33}X_3 + A_{43}D_4 = (0.90)(4000)+(0.70)(800)+(0.90)(400) + (1.00)(800) = 5320 \text{ vph} > B_3 = 5200 \text{ vph}; \therefore \underline{X_4 = 680 \text{ vph}}$
- $S_4 = A_{14}X_1 + A_{24}X_2 + A_{34}X_3 + A_{44}X_4 + A_{54}D_5 = (0.85)(4000) + (0.60)(800) + (0.85)(400) + (0.90)(680) + (1.00)(600) = 5432 \text{ vph} > B_4 = 5200 \text{ vph}; \therefore \underline{X_5 = 368 \text{ vph}}$

<u>Conclusion:</u>

- Ramp #1: No control needed.
- Ramp #2: Meter at a rate of 400 vph.
- Ramp #3: Meter at a rate of 680 vph.
- Ramp #4: Meter at a rate of 368 vph.

Figure 7.6. Integrated Entrance Ramp Control Example Number 1: Calculation of Pretimed Metering Rates

Example Number 2 — The data given in the example shown in figure 7.7 is the same as that given in the previous example, except that the mainline demand D_1 is 4,600 vph instead of 4,000 vph. In this case, the metering rates at ramps 2, 3, and 4 are determined solely by their respective downstream available capacity constraints, as was the case in the previous example. However, the metering rate at ramp 1, rather than being determined by the available capacity constraint at section 1, is established in accordance with the available capacity constraint at ramp 2, as is described below.

The demand S_2 at section 2 is 5,570 vph, which is 770 vph greater than the capacity B_2 at section 2 (4,800 vph). If ramp 2 is closed, the demand at section 2 is reduced to 4,970 vph, a volume which also exceeds the capacity B_2. Therefore, it is necessary to reduce the allowable volume X_2 entering at ramp 1 (input 2). The allowable volume X_2 must be reduced enough to reduce the demands S_2 by 170 vph. The amount of the reduction is equal to the 170 vph divided by decimal fraction A_{22} of the vehicles entering at ramp 1 and passing through section 2 (170 vph/0.75 = 227 vph). Therefore, the allowable volume X_2 at ramp 1 would be 573 vph instead of 800 vph.

In the procedure outlined above, excess demand $S_j - B_j$ at any section j is removed by reducing the allowable volume on the entrance ramp immediately upstream. If, instead, the allowable volumes on entrance ramps farther upstream were reduced, a larger number of vehicles would have to be removed from these ramps in order to reduce the demand S_j sufficiently at section j. This is necessary because some of the vehicles that enter these ramps will exit the freeway before they reach section j.

Example Number 3 — Considering again the situation presented in figure 7.6, allowable ramp volumes would be calculated as follows. If the excess demand (200 vph) at section 2 were to be removed by reducing the allowable volume X_2 at ramp 1, the volume at ramp 1 would have to be reduced by 267 vph. Consequently, the allowable entrance ramp volumes would be the following:

	Volume (vehicles per hour)
Ramp Number 1	533 vph
Ramp Number 2	600 vph
Ramp Number 3	687 vph
Ramp Number 4	352 vph
Total Input	2,172 vph

The total input of 2,172 vph, however, is less than that of 2,248 vph, the volume which is obtained if ramp 2 is metered as in Example Number 1. Thus, the fundamental approach described will result in the optimal utilization of the freeway. It maximizes the sum of the allowable entrance ramp volumes, a procedure that corresponds to maximizing system output for steady-state, uncongested flow conditions. Total travel in the system is also maximized.

Linear Programming Formulation

The fundamental procedure described for examples 1 and 2, has been formulated as a linear programming model(4), which is used to compute optimal allowable entrance ramp volumes. In terms of the notation defined in figures 7.6 and 7.7, the linear programming model would be:

Step 1: Maximize ΣX_i, where n is the number of inputs.

Step 2: Subject to the following constraints.

240 vph at ramp 2 in the example presented in figure 7.7, it would be necessary to follow the adjustment procedure for the metering rates at the other entrance ramps (as shown in figure 7.8). The allowable volume X_2 (573 vph) at ramp 1 would have to be reduced by 320 vph in order to allow 240 vph to enter ramp 2 and still satisfy the available capacity constraint at section 2. This reduction also decreases the mainline demand at sections 3 and 4. Thus, the allowable volumes at ramps 3 and 4 are increased to maximize the utilization of the freeway at these sections.

Integrated Traffic-Responsive Metering

Integrated traffic-responsive metering is the application of traffic-responsive metering to a series of entrance ramps where metering rates at each ramp are selected in accordance with system, as well as local, available capacity constraints.

During each control interval, real-time measurements are taken of traffic variables (usually volume, occupancy, and/or speed). The data are used to define the available capacity conditions at each entrance ramp. Then, on the basis of these measurements, both an independent and an integrated metering rate are calculated for each entrance ramp. Of these two metering rates, the one that is the more restrictive is selected to be used during the next successive control interval.

The methods used to calculate the independent and integrated traffic-responsive metering rates are basically the same as those used to compute independent and integrated pretimed metering rates. Usually, rather than calculating metering rates in real time, a set of metering rates is precomputed for the range of expected available capacity conditions from which metering rates are selected in real time. The linear programming model is often used to calculate predetermined sets of integrated traffic-responsive metering rates. Also, the metering rates are usually subject to the overrides for the merge detector, queue detector, and maximum red time used in traffic-responsive metering.

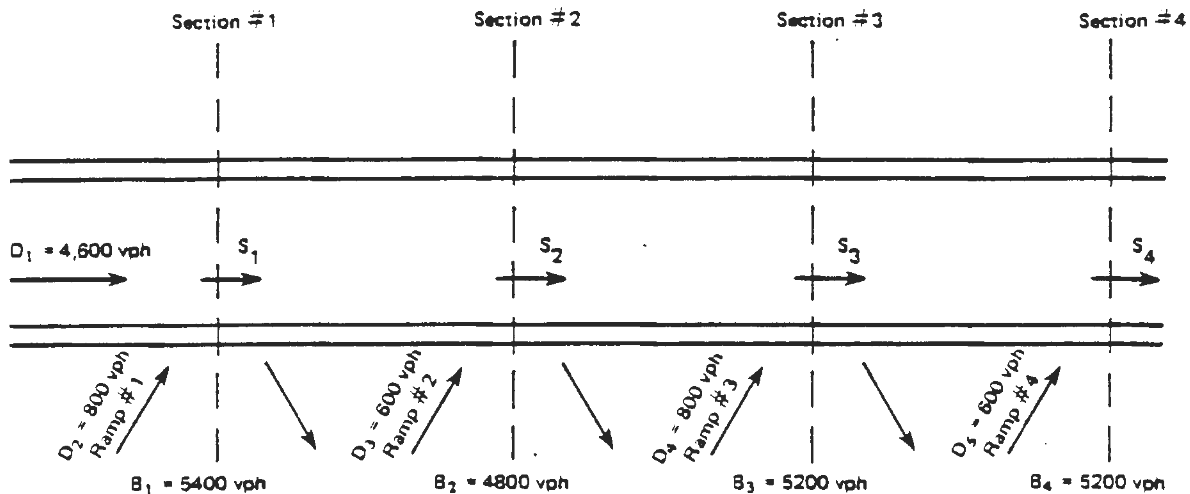
Integrated Gap Acceptance Merge Control

Integrated gap acceptance merge control is the application of gap acceptance merge control to a series of entrance ramps where the gap acceptance merge control at the individual ramps is subject to system available capacity constraints.

Metering rates are computed by the procedure described for calculating integrated pretimed metering rates. Integrated metering rates are computed for each entrance ramp on the basis of the real-time measurements of traffic variables that are used to define the system's available capacity conditions. Then, for each entrance ramp, a minimum acceptable gap setting is determined, which would yield a ramp volume that corresponded to the integrated metering rate.

The minimum acceptable gap is the smallest gap in freeway traffic into which a ramp vehicle will merge. The larger the minimum acceptable gap, the lower the resulting ramp volume. The smaller the minimum acceptable gap, the higher the resulting ramp volume.

The evaluation of differing control strategies and ramp metering rates for large systems is best performed by simulation models. These models are described in chapter 3 of this volume.



A_{ij} Values

i \ j	1	2	3	4
1	1.00	0.95	0.90	0.85
2	1.00	0.75	0.70	0.60
3	–	1.00	0.90	0.85
4	–	–	1.00	0.90
5	–	–	–	1.00

X_i = allowable volume at input i

D_i = demand at input i

B_j = capacity at section j

A_{ij} = decimal fraction of vehicles entering at input i which pass through section j

S_j = demand at section j

Compute X_i's starting at section #1

- Set $X_1 = D_1 = \underline{4600 \text{ vph}}$
- $S_1 = A_{11}X_1 + A_{21}D_2 = (1.00)(4600)+(1.00)(800) = 5400$ vph $= B_1 = 5400$ vph; $\therefore X_2 = \underline{800 \text{ vph}}$
- $S_2 = A_{12}X_1 + A_{22}X_2 + A_{32}D_3 = (0.95)(4600)+(0.75)(800)+(1.00)(600) = 5570$ vph $> B_2 = 4800$ vph;

  Since X_3 must equal at least 240 vph; $X_3 = \underline{240 \text{ vph}}$ and volume entering upstream must be reduced by 410 vph;
  $\therefore X_2 = 800 - 410/A_{22} = 800 - 410/0.75 = \underline{253 \text{ vph}}$
- $S_3 = A_{13}X_1 + A_{23}X_2 + A_{33}X_3 + A_{43}D_4 = (0.90)(4600)+(0.70)(253)+(0.90)(240)$
  $+ (1.00)(800) = 5333$ vph $> B_3 = 5200$ vph; $\therefore \underline{X_4 = 667 \text{ vph}}$
- $S_4 = A_{14}X_1 + A_{24}X_2 + A_{34}X_3 + A_{44}X_4 + A_{54}D_5 = (0.85)(4600) + (0.60)(253) + (0.85)(240) + (0.90)(667)$
  $+ (1.00)(600) = 5468$ vph $> B_4 = 5200$ vph; $\therefore \underline{X_5 = 334 \text{ vph}}$

Conclusion:

- Ramp #1: Meter at a rate of 253 vph.
- Ramp #2: Meter at a rate of 240 vph.
- Ramp #3: Meter at a rate of 667 vph.
- Ramp #4: Meter at a rate of 334 vph.

Figure 7.8. Integrated Entrance Ramp Control Example Number 4: Calculation of Pretimed Metering Rates

REFERENCES

1. FHWA. "Traffic Control Systems Handbook." January 1977.

2. ITE Committee 4Z-A. "Freeway Entrance Ramp Displays." Tentative ITE Recommended Practice. September 1974.

3. Everall, P.F. "Urban Freeway Surveillance and Control: The State of the Art." FHWA. November 1972.

4. Berry, J.A., and Wattleworth, D.S. "Peak-Period Control of a Freeway System — Some Theoretical Investigations." Highway Research Record No. 89.

CHAPTER 8. MAINLINE AND CORRIDOR CONTROL

INTRODUCTION

Mainline control is concerned with the guidance, regulation and warning of traffic on the freeway—not on the entrance ramps. Corridor control is designed to better utilize all available facilities within a corridor that includes the freeway and adjacent parallel arterial roads. Surveillance of the freeway, its entrance ramps, and parallel arterial roads provides information enabling the control of demand on the freeway when congestion occurs or is imminent following excessive demand or an incident.

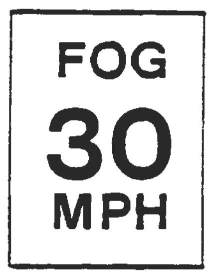
Due to the interdependence between mainline and corridor control, this chapter discusses both methods in parallel.

Mainline and corridor control, like other forms of traffic control, are aimed at increasing the efficiency and safety of traffic flow. Particular objectives are to:

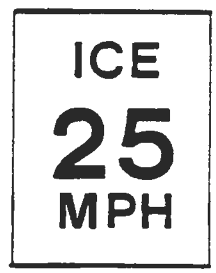
- Reduce rear-end collisions by informing drivers of congestion ahead
- Improve the stability of traffic flow so that reduced variance in freeway speeds causes the facility to be more efficient and lessens the likelihood of accidents
- Facilitate incident management and recovery
- Reduce driver impatience and frustration by providing reasons for delay
- Divert traffic to alternate routes when congestion occurs
- Change the directional capacity of the freeway by using reversible lanes

MAINLINE SPEED CONTROL

Mainline speed control is a technique that uses signs to limit the speed of traffic and reduce the likelihood of rear-end collisions. This technique has been less successful when a reason for the speed reduction was not provided. It was found that drivers did not readily alter speeds, and considered the posted signs advisory, instead of regulatory. Experience in Europe has shown that when speed regulation signs are used in conjunction with driver information signs, they are effective and produce a significant reduction in speed. The European practice is to use a message sign in conjunction with a speed sign such as:



or



This technique has been found to reduce rear-end collisions by reducing both vehicular speed and the standard deviation of speed.

DRIVER INFORMATION SYSTEMS

There are several basic principles that apply to signing used to provide driver information.

Emphasize Important Information

One of the basic tenents of sign design is that the most important information should provide the most impact. Signs for principal routes (such as freeways and interstate roads) should be more dominant than those for smaller roads with lesser demand. The letter size of a sign at an off-ramp should be larger than the letter size of a similar sign further upstream.

Avoid Providing Too Much Data

Information overloading should be avoided. If too much data is contained in a sign, its effectiveness is reduced. A long reading time caused by excessive information also has safety implications. During the time taken to read a sign, drivers are distracted from watching traffic, and this has the potential of increasing accidents.

Provide Advance Information

Providing information in advance is of particular importance and, where possible, the distance to the interchange should be included with allowance made for the number of lanes and typical flow levels that drivers have to cross.

Distribute Signing

Drivers on freeways, particularly those on long trips, are subject to fatigue and boredom. This can result in accidents. By distributing signing and spreading out information, individual drivers are kept informed and interested.

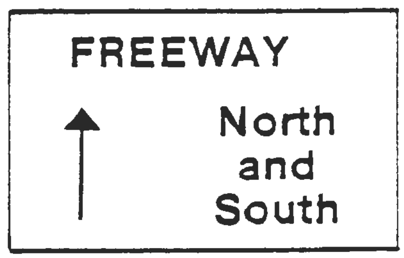
The careful placement of signs by the engineer and avoidance of providing unexpected information can be beneficial to the driver and the manager of the freeway system.

There are two types of signs used in driver information systems:

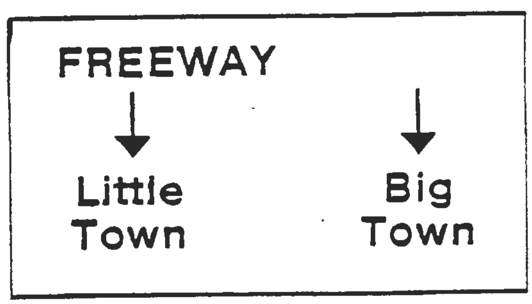
- Fixed signs
- Variable signs

Fixed Signs – When considering freeway signing, engineers should consult the "Manual on Uniform Traffic Control Devices"(1). This manual provides detailed guidance on the location and design of fixed signs. Further information on sign construction and costs is contained in volume 3 of this handbook.

Unfortunately, there are many examples of signs that conform to the manual in lettering size and background, but are ineffective because they were poorly designed. Examples exist of badly placed signs and signs that are obscured by vegetation or bridges. The principal purpose of fixed signs is to direct unfamiliar drivers. A common error is to assume a prior knowledge on the part of the driver. An example of this assumption would be:



and further down the road the next sign:



An unfamiliar driver who has no direct knowledge of the area, but knows that he wants to go north is confused in situations where fixed signs such as these are used.

Variable Signs — Variable signs include:

- Single-message signs with a fixed sign face that flashes when a particular hazard occurs
- Secret signs that are blank unless activated
- Variable message signs that can provide a variety of information

In the context of corridor and mainline control, all of these message types can be useful in providing information to the driver and improving traffic flow. Single-message signs and secret signs provide one message at a time to give advance warning of conditions that change over time such as:

- SLOW — CONGESTION AHEAD
- SLOW — ICE
- ROAD CLEAR

Variable message signs give the enigneer an opportunity to provide a range of information to the driver. Details of the types of variable message signs are contained in volume 3. Figure 8.1 shows a sign that provides a variety of messages. The location of signs is an important factor that the engineer should consider when planning a system. If diversionary information is provided, drivers should have the capability of exiting. For example, if a sign indicates congestion downstream, the sign should be located upstream of an exit ramp. When such a sign is activated, allowance should be made, if possible, to facilitate the additional traffic on arterial streets.

Variable message signs can also be used on the arterials leading to freeway ramps in order to inform drivers in advance of freeway conditions. Again, the location and sign content are important. Such signs should be placed so the driver can:

- Divert to another ramp, if desired
- Decide not to use the freeway at all, and select an arterial route

An example of an off-the-freeway sign is shown in figure 8.2. Some sign designs make effective use of color by indicating congested traffic or downstream delay times in red and recommended downstream entrance ramps in green. Colors do have an extra attention-getting capability, and drivers often associate red and green with recommendations to stop or go.

The shoulders and median of a freeway present a particularly hostile environment. Any sign adjacent to freeway traffic is not only subject to extreme temperatures and precipitation, but also to considerable amounts of road dust and splashing water from passing traffic. The long-term reliability of some changeable message signs that use mechanical devices should be questioned closely. Matrix signs with no moving parts are considerably less expensive than "roller blind" mechanical signs, and have been found to be more reliable in adverse weather conditions. This is particularly important as variable message signs are often used during adverse weather conditions.

CORRIDOR DIVERSION

When considering the flow of traffic in a corridor that contains both a freeway and an arterial road network, it is desirable that as much traffic as possible use the

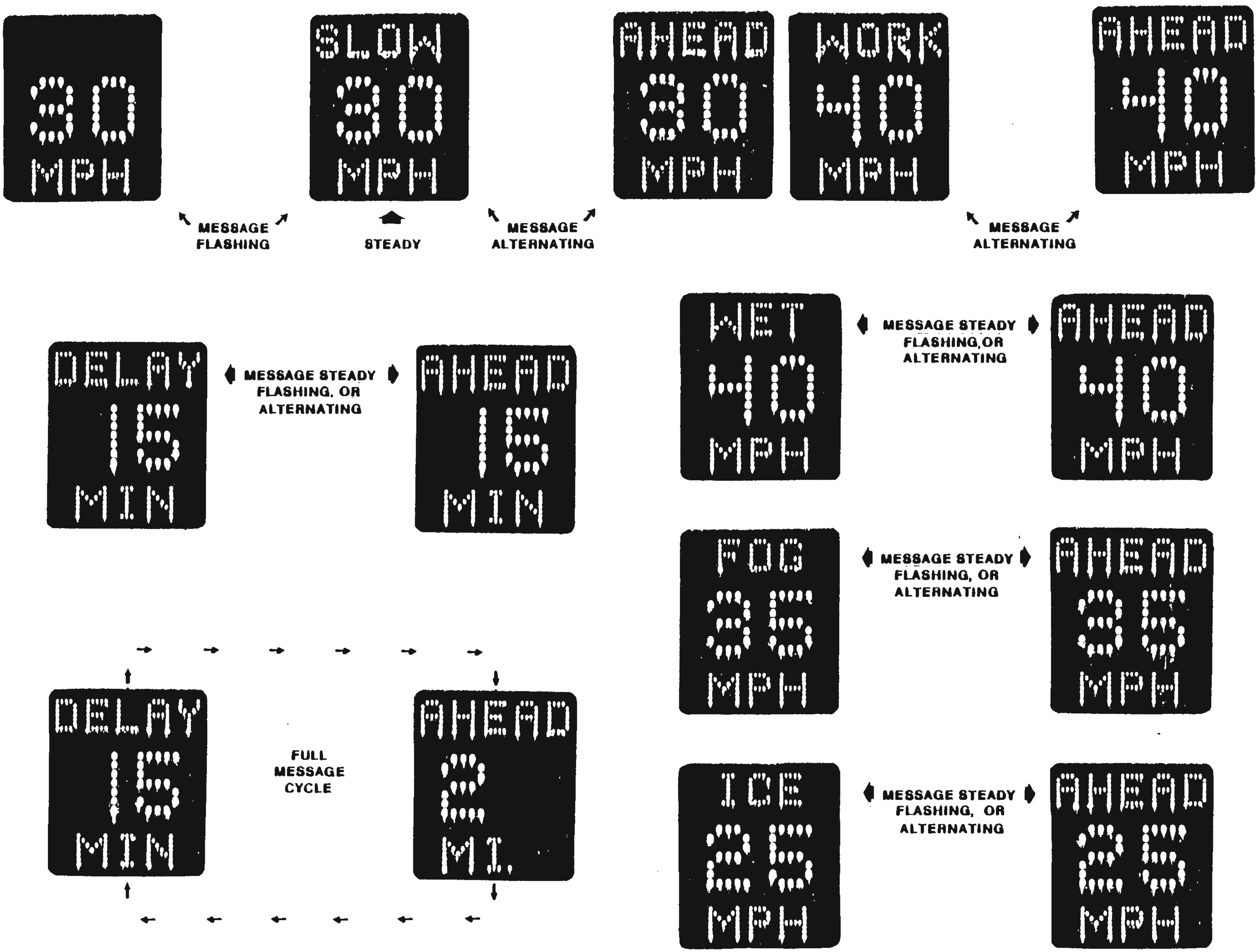


Figure 8.1. On-Freeway Advisory Signs

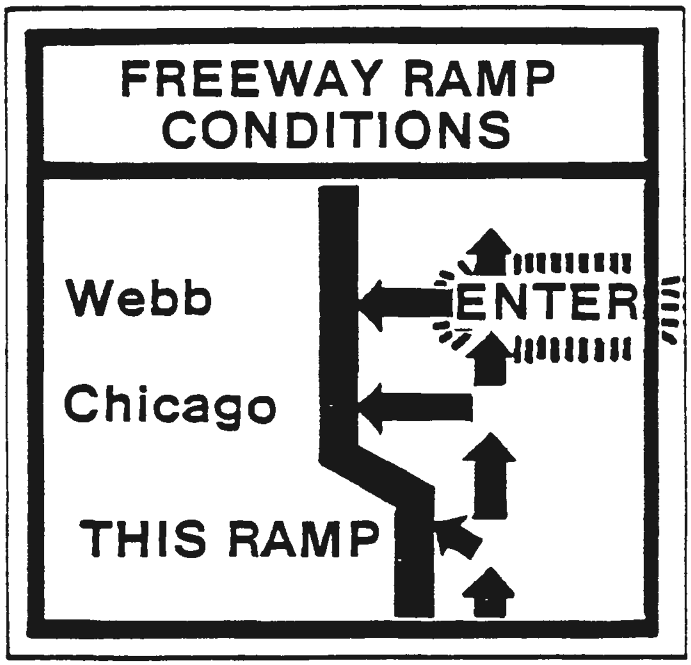


Figure 8.2. Example from Detroit of an Off the Freeway Variable Message Sign

freeway until the delay to freeway traffic becomes comparable with the delay on the arterial network. Compared with driving on a freeway, arterial street traffic is subject to lower speeds, intersection delay, and increases in noise and pollution. When the freeway is approaching capacity, however, the overall operation of the corridor can be enhanced by the application of diversion techniques. Figure 8.3 shows a section of a typical urban corridor in which demand exceeds capacity (recurring congestion), incidents occur (non-recurring congestion), and maintenance activities are performed.

Diversion During Recurring Congestion

Usually the first reason for demand to exceed capacity can be predicted. This is discussed in chapter 6 of this volume, "Detection of Recurring Congestion." By informing drivers already on the freeway, as well as those intending to use the freeway, that delays should be expected, some traffic will divert to arterial roads and relieve the situation on the freeway. The use of corridor diversion techniques for situations that can be foreseen (recurring congestion) does not necessarily require that the arterial intersections are linked under some form of central control. Since freeway overcapacity is often influenced by time of day, the identification of peak periods can aid in setting the phases of the arterial intersection controllers. This retiming will optimize the use of the arterials.

Diversion During Non-Recurring Congestion

For incidents occurring on the freeway that cannot be predicted, having the arterial traffic signals centrally coordinated provides the engineer with the ability to make the most efficient use of all roads in the corridor. By hypothesizing incidents on each freeway segment and evaluating feasible diversion routes, the engineer can have prepared signal timing plans that allow for the additional flow resulting from the diversion.

Road Maintenance

Maintenance activities fall somewhere between the last two cases. Since short-term maintenance schedules are generally known, appropriate temporary diversion signs or necessary adjustments to arterial signal timing plans can be used to minimize effects of freeway maintenance on traffic flow.

LANE CLOSURE

The two principal reasons for closing a freeway traffic lane are maintenance and incidents. (Techniques for lane closure during incidents are discussed in volume 3.) In Detroit a technique was used to signal lane closure during maintenance operations by using a sign above the freeway showing a vertical green arrow that was changed to a red X when the lane was closed. The effectiveness of this technique was found to be a function of demand. When the demand was less than the capacity at the construction region, drivers obeyed this sign. However, when the demand was greater than the capacity in the construction area, there appeared to be no increase in the capacity of the bottleneck. No accident data were recorded in this study, which is unfortunate as this could have been proved to be a useful technique. Research in Europe has shown that the use of lane closure signs is effective in reducing the frequency and severity of accidents. Recently, the Transport and Road Research Laboratory has developed a matrix sign for use on U.K. motorways. The sign consists of three segments:

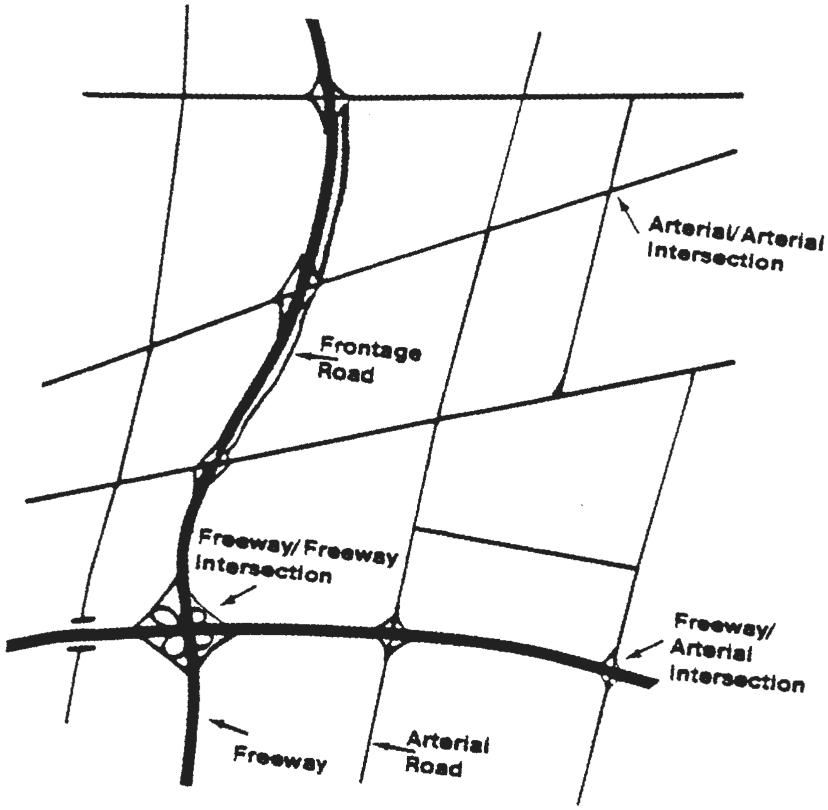


Figure 8.3. A Freeway Corridor

- Top—Indicates lane closure or speed
- Middle—Displays a pictogram of the reason for warning
- Bottom—Shows the distance to the hazard

Figure 8.4 shows the three-part sign and figure 8.5 shows the proposed pictograms that would be used in the middle of the sign.

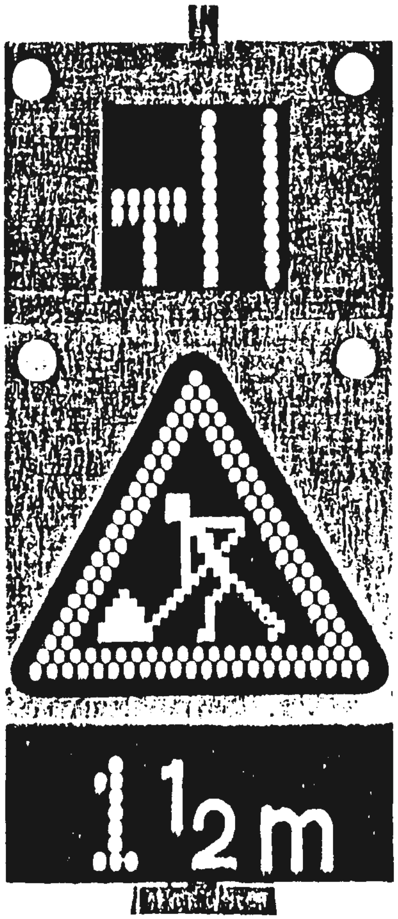


Figure 8.4. Three-Part Signal Used in the Trail, Showing a Typical Legend Combination

Figure 8.5. Proposed Pictograms for the Enhanced Motorway Signal

REFERENCES

1. FHWA. "Manual on Uniform Traffic Control Devices." 1978.

CHAPTER 9. BUS/CARPOOL PRIORITY CONTROL

INTRODUCTION

Modern urban families with a desire for individual housing have led to a massive growth of suburban areas and an increasing dependency on the automobile. The car has several distinct advantages when compared to mass transit systems:

- Door-to-door service
- The car can be used on an individual's timetable, not one specified by a transit company
- Protection from the environment—no waiting in the cold or rain
- Reliable service
- Shorter traveltimes
- Greater comfort

Carpooling, although overcoming some of these problems, does have disadvantages to the average commuter:

- Lack of commuters with similar origins and destinations
- Lack of information-gathering services to discover similar commuters
- Non-flexible timetable

With all the advantages of the private car over the use of buses or carpools, how can the engineer persuade drivers to use high-occupancy vehicle (HOV) techniques? There are two basic approaches. The first is to make attempts to improve the quality of service provided by HOV techniques, the carpool and vanpool ride-sharing programs, preferential parking for carpools and vans, making timetables and routes optimum for the user, providing shelters and seats at stops, and making buses more comfortable. The second approach is to emphasize the commuter's savings in time and money which can be made by using carpools, vanpools, and buses on HOV facilities.

The most successful HOV facilities are those with large time savings. A finding(1) in this regard appears to be that use of HOV facilities becomes significant as time savings perceived by motorists exceed 1 minute per mile.

Care should be taken in the planning stage to ensure a reasonable and feasible HOV project design. HOV lane projects that have aroused the greatest public opposition (and eventually were terminated) have often shared the characteristic of preempting an existing peak direction travel lane for HOV use. The reduction in highway capacity and ensuing increased congestion aroused motorist resentment. Public reactions were often reflected in the media, which aroused further adverse public sentiment.

Enforcement, unpopularity, apparent lack of full HOV lane utilization, and weak political support, have been cited as added reasons for failure of the concurrent-flow/existing lane projects.

HOV lane projects that have been more successfully received by the public are those where HOV lanes were added, thereby increasing capacity; HOV lane utilization was high; police enforcement was effective; and an extensive marketing effort preceded implementation of the HOV lane.

DESIGN

Classification

The usual classification scheme applied to HOV lane projects is centered on the physical configuration of the HOV lane relative to the freeway in which it has been incorporated:

- Separated — HOV lanes are physically separated from regular lanes by a median, separation, or other barrier.

- Contraflow — HOV lanes are preempted from the regular lanes, having traffic flowing in the direction opposite the predominant peak-flow direction (e.g., an outbound lane in the a.m. peak when the heaviest flow is inbound). The use of the lane as an HOV lane may be indicated by signing, signals, striping, pavement symbols and cones, or similar means.

- Contraflow/Concurrent-Flow — HOV lanes are preempted from regular lanes carrying traffic in both directions in the peak hour, producing two lanes bearing HOV's traveling in the peak-flow direction.

- Concurrent-Flow/Lane Added — An HOV lane is constructed alongside existing regular lanes and so designated by appropriate signing, striping, coning, or other evidence of its intended use and the hours when such use applies. Absence of physical separation from regular freeway lanes facilitates use of the HOV lane for regular traffic in off-peak hours.

- Concurrent-Flow/Existing Lane — In this configuration, the HOV lane is preempted from the existing regular lanes bearing traffic traveling in the peak-flow direction.

Figure 9.1 illustrates these HOV configurations. The layouts of lanes shown are not all-inclusive as differing configurations do exist, but these are the most common.

The designation of which vehicles use HOV facilities varies with differing installation; some are restricted to buses, others allow buses and carpools. Usually carpools are restricted to three or four occupants. When evaluating plans concerning potential usage of HOV facilities, the engineer should make allowance for future increases in traffic flow. To allow buses and carpools today, and then to limit the facility to allow buses only in the future causes an adverse public reaction. Similar bad reactions have been found when changing carpool occupancies from three to four persons. There seems to be a good argument for making the restrictions harsh at the inception of a project and, if it is deemed necessary following a review of system performance, the restrictions can be reduced. HOV facilities that appear underutilized bring an adverse reaction from occupants of the congested regular lanes who question the justification of the HOV facility.

Lack of flow on HOV lanes was given as a major reason for the public outcry that led to the termination of Boston's Southeast Expressway project. It also led to a severe truncation of the hours of operation on the Banfield Freeway in Portland.

Generally, the longer HOV lanes are more successful in meeting the goals of mass transit systems. Longer HOV lanes produce larger gains in time, and this influences motorists to consider a change in their mode of travel.

Hours of operation for all HOV lane projects have been limited to the 3 to 4 peak hours in the morning and evening. Some projects that originally set the period of

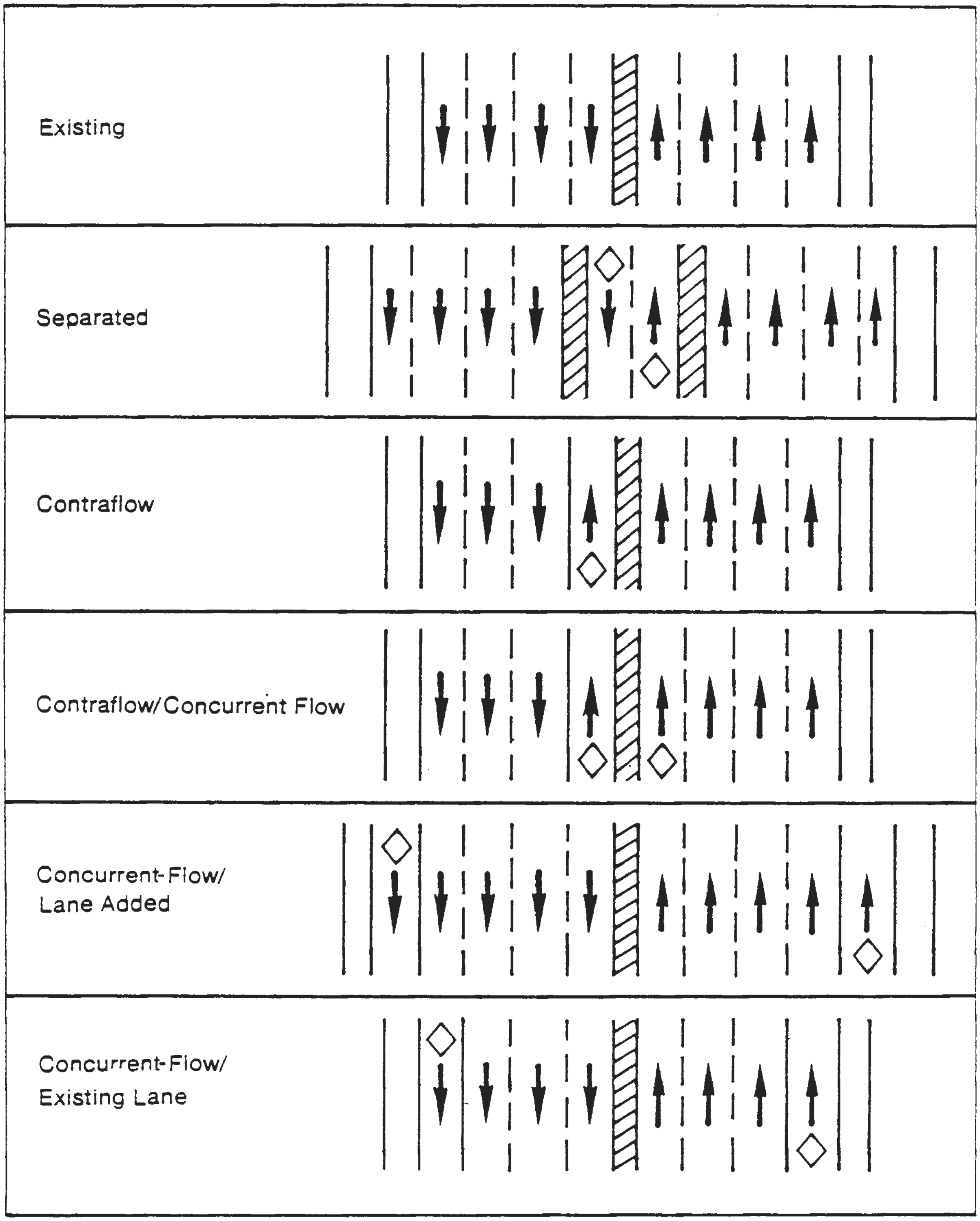


Figure 9.1. HOV Lane Classifications
(◇ Denotes HOV Lane)

operation at 24 hours were influenced by adverse public opinion to narrowly restrict HOV lane operation to peak periods. In one instance, safety considerations were involved: in the off-peak hours the empty and specially striped HOV lane was perceived by some as a breakdown lane, and accidents began to occur as disabled vehicles parked in the HOV lane to effect emergency repairs. As a practical matter, there appears little reason to restrict general traffic from the HOV lane in off-peak hours. The only problem reported from the limited operating hours is that violation rates tend to be higher during the transition periods (the quarter-hour at the beginning and end of the HOV lane-operating period) when some drivers may not be aware of the precise time. The signs associated with HOV facilities are discussed in volume 3.

Person Flow

There are many measures of effectiveness that can be used when evaluating a freeway with an HOV facility. One of the most relevant measures is person flow. An effective HOV lane will carry a greater number of persons than a regularly traveled lane. For example, consider a four-lane freeway that is being considered for conversion to a concurrent-flow/existing lane HOV facility. Before the conversion, the lane under examination would have carried 25 percent of the persons. If after studies showed no increase in this figure, the costs associated with HOV conversion would be money wasted, unless there are other local factors affecting the evaluation. For those types of HOV facilities where an additional lane is provided, the person-flow figure can be compared with how high the flow would be if the lane were traveled by all types of traffic. Table 9.1 shows the percentage flow of persons if the lane were traveled by all traffic, and the actual person flow.

Priority Entry

Priority entry is a technique by which high-occupancy vehicles are given access to a facility via special lanes. These special lanes might be exclusive bus ramps, bypass lanes at ramp meters, or special reserved lanes at a toll booth. This technique can be applicable to older freeways where right-of-way is limited, volumes heavy, and little potential exists for exclusive lanes, contraflow lanes, or concurrent-flow lanes. Priority entry can be implemented at relatively moderate cost, and provide an increase in people-moving capacity with little negative impact.

Priority entry techniques can also be used in conjunction with HOV lanes. In two experiments—one in Los Angeles and one in San Diego—the application of this technique resulted in more than doubling the number of carpools. Some of the advantages of using priority entry techniques include:

- The underutilization of a freeway lane due to insufficient bus patronage and carpoolers is avoided as the vehicles join the general stream of traffic.
- The safety problems associated with the speed differential between priority and other vehicles is removed.
- Enforcement is easier on the ramp where speeds are low and violators can be readily identified.
- Capital and operating costs are low.

The disadvantages include:

- Those vehicles given priority on entrance ramps are subject to the same congestion as the general mainline traffic and thus the overall traveltime savings are small.

TABLE 9.1. PERSON-FLOW PERCENTAGES

	Percent of Peak-Period Person Throughput If Regular Lane	Percent of Peak-Period Person Throughput Actually In Priority Lane	Difference	Minutes Saved
Shirley Highway	40	68	+28	15 to 20
San Bernardino Freeway	20	23	+3	18/8
I-93 Boston	33	21	-12	4
I-495/Lincoln Tunnel	25	N.A.*	N.A.	8 to 15
Long Island Expressway	25	N.A.	N.A.	15
Boston SE Expressway	33	N.A.	N.A.	14
U.S. 101 Marin County	40	39	-1	1
I-95 Miami	20	20	0	2
Banfield Freeway	33	19	-14	1
Moanalua	25	19	-6	5
Santa Monica Freeway	25	14	-11	6
Boston SE Expressway	50	40	-10	5

* Data were not available on the bus-only contraflow projects. Table 9.1 also shows the minutes saved to travelers using the HOV facility. It can be seen that the projects with a high percentage use of the HOV facility also have a high time saving. The first six projects in the table are separated or contraflow projects indicating that these are generally more successful than the alternative designs.

- In some circumstances, the provision of bypass lanes on ramps could be impossible due to the proximity of adjacent roads and buildings.

HOV Lanes Upstream of Bottlenecks

Although the decision to install an HOV facility should be made in terms of person flow and traveltime savings, in the situation where bottlenecks occur, an HOV facility of short length can be justified. Obviously, where bottlenecks occur, the best solution is to eliminate the cause of congestion. However, this is often unfeasible or too expensive. In these circumstances, the engineer would consider an exclusive lane facility only over the length of the maximum queue usually found upstream of the bottleneck. An inexpensive technique for achieving this is to use the existing facility's shoulders. The feasiblity of this technique must be carefully analyzed from a safety and enforcement standpoint. Advantages of utilizing exclusive lanes upstream of bottlenecks would include:

- Low operating costs
- Reduced traveltime for persons in the HOV lane
- If shoulders or an additional lane are used, other traffic is not adversely affected

There are some potential problems associated with short HOV lane lengths upstream from bottlenecks. These include:

- Safety problems could be encountered at the merging area due to HOV's that are traveling quickly, merging with slower moving regular vehicles.
- Enforcement could be difficult if the shoulder is used as a traveled lane, since police officers would have no location in which to stop and cite violators.

ENFORCEMENT

A number of HOV projects have experienced sub-optimal levels of enforcement. This is due in part to a lack of engineering concern with enforcement, even though the enforcement issue has a considerable impact on the operational and safety characteristics of HOV projects, especially those where significant modifications to existing traffic patterns occur. As diversification in the design of HOV projects continues, the issue of enforcement of HOV facilities takes on greater importance, and the need to develop enforcement strategies that involve a systematic approach to violator apprehension becomes essential(2).

Traffic law enforcement personnel should be involved in the planning of an HOV facility. The effort to obtain insight into the nature of possible enforcement problems and to gain the support of law enforcement agencies can provide the transportation engineer with valuable information. In many cases, compromises may have to be made in terms of the final design concept and/or the desired enforcement program.

In selecting a final HOV design strategy for implementation, the enforceability of that concept should be taken into consideration. For each HOV design strategy, the project-planning and design team should question the difficulty involved in enforcing the restrictions associated with each proposed strategy. Possible modifications to HOV design strategies should be explored to alleviate as many potential enforcement problems as possible during the planning and design stage.

Once the HOV design concept has been selected from a number of candidate strategies, a comprehensive enforcement program should be developed. It is possible that several enforcement strategies, or more specifically several sets of procedures within a given strategy, may be applicable to the realistic enforcement objectives of any given HOV project. A careful review of the local legal environment and state statutory requirements should be made, particularly if innovative enforcement practices are under consideration. There are two basic criteria that can be used to judge the performance of various enforcement options: the projected violation rate and the projected cost of the enforcement program. The selection of the alternative that produces the best results per dollar invested can be made in a straightforward manner. Unfortunately, there is a lack of detailed statistical information needed to provide a scientific process for forecasting the violation rate.

In view of the lack of precise data on which to base the design of the final enforcement program, it is recommended that an evaluation plan be developed to assure a continuing flow of empirical data and feedback for program optimization. Specific areas relating to HOV lane operations and enforcement operations that should be quantified include:

- Relationship between the number of citations issued and the number of violations occurring
- Interrelationships between the violation rate, apprehension rate, and the traveltime savings of the HOV lane
- Changes in the violation rate due to changes in the quantitative, qualitative, or substantive aspects of the enforcement program

It may be possible to reduce the enforcement level of effort without compromising HOV lane operations and enforcement objectives.

A detailed enforcement manual is highly recommended for effectively managing a complex HOV program. This manual should provide descriptions on the HOV project, system operations, enforcement procedures, and reference information. A sample enforcement manual is provided in reference 3.

Public awareness is essential in any new enforcement program. If the public is made to understand the HOV operating strategy and its restrictions, the tendency to violate may be reduced. Furthermore, enforcement agencies uniformly concur that a public awareness program that notifies the public of enforcement activities increases the effectiveness of the enforcement effort. Inexpensive public education techniques available include news releases and conferences, public service advertising, transit advertising space, speakers' bureaus, pamphlets or handouts, and banners over the roadway. More expensive techniques include paid TV, radio, and newspaper advertising, as well as roadside billboards. The primary message transmitted with respect to HOV enforcement education should be a simple statement of what the law prohibits and the consequences of violating that law.

From the research conducted on various HOV projects, transportation and enforcement officials have identified a number of problems associated with their HOV project enforcement programs. The problems are created by geometric, operational, or institutional factors, and include:

- Lack of a safe and easily accessible refuge area bordering the HOV lane, which can be used to apprehend and cite HOV violators
- Absence of any vantage point by which enforcement officials can observe the HOV facility while keeping out of view

- Concurrent-flow HOV projects that do not physically separate the HOV lane from the general traffic lanes, thereby providing the motorist with an infinite number of locations at which HOV regulations can be violated

- HOV facilities lacking a paved surface, clear of obstructions, for passing

- Determination of the number of occupants in a vehicle is made difficult by (1) young children, (2) vans, mobile homes, etc., (3) mirrored glass, (4) hours of darkness, and (5) inclement weather

- HOV projects having a speed differential between the HOV lane and the general traffic lanes that presents a significant safety concern for all traffic and especially enforcement personnel

- For HOV projects where refuge areas are not adjacent to the HOV lane, the citing of HOV violators is less visible to the motorists

- HOV restrictions requiring judgment decisions on the part of the enforcement personnel, primarily for curb bus lanes and the use of the bus lane by right-turning vehicles

- Proper coordination and cooperation between project management, enforcement, and judicial interests

- Geometric or operational problems that make it extremely difficult for the witnessing officer to be the apprehending officer

- Manpower constraints

- Little incentive toward compliance with HOV restrictions, caused by a low probability of being cited, and/or a low fine

<u>Enforcement Effectiveness</u>

The primary measure of effectiveness of an HOV enforcement program is the violation rate achieved. On most projects, the violation rate is defined as the percent of the total number of vehicles using the HOV lane which fail to meet eligibility criteria for the HOV lane. The violation rates for the HOV projects encompass a wide range of percentages—from nearly a zero percent violation rate to a violation rate of over 50 percent. Other projects with similar geometry and operating strategies can have drastically different violation rates because of the type and level of enforcement employed.

The fact that an HOV project is experiencing a relatively high violation rate may not necessarily indicate failure of the HOV project objectives. The intent of employing a certain type of enforcement strategy is, in part, to achieve a violation rate that is agreed upon as tolerable to project management, enforcement personnel, motorists, and the general public. A high violation rate could very well be considered to be tolerable by the policy-making group.

There are a number of factors that affect the violation rate. These include:

- HOV lane signing

- Bus versus carpool lane restrictions

- Traveltime savings

- Probability of apprehension

- Accessibility to the HOV lane

- Hours of operation

- Occupancy restrictions

- Visibility
- Weather conditions

When a motorist willfully violates the HOV lane, he often believes that he has a very good chance of escaping apprehension. In short, the motorist's perceived benefits outweigh the perceived risks. The overwhelming benefit that a motorist would receive is the traveltime savings in the free-flowing HOV lane, as opposed to the congested general lanes. However, if the probability of being apprehended for the HOV violation is 100 percent, then the violation rate would approach zero regardless of the magnitude of the traveltime savings. The probability of being apprehended and cited for an HOV violation is dependent upon:

- The number of enforcement personnel
- The time taken for detection, apprehension, and citation procedures
- The length of the facility
- The number of HOV violators

Generally, one of the objectives of HOV projects is to improve traffic flow on the particular facility. However, enforcement of the HOV projects often disrupts traffic flow. The directly related traffic flow problems are mainly associated with an apprehension procedure resulting in hazardous weaving maneuvers performed by the enforcement vehicle and/or the violator's vehicle. Once an HOV violator is escorted to a refuge area, the enforcement effort can be indirectly involved in disrupting traffic flow and contributing to traffic accidents through "rubbernecking."

There are certain recommendations on enforcement of HOV priority treatment projects that are common to all freeway applications. These recommendations are:

- Enforcement requirements should be included in project planning in the earliest stages, and enforcement personnel should be active members of the planning team.
- To the maximum extent possible, HOV priority projects should be designed, constructed, and/or modified in strict conformance to AASHTO and MUTCD standards.
- Officials of the traffic court system should be briefed, prior to project startup, regarding the project's operational goals, traffic restrictions, enforcement program, and legal basis.
- On projects having traveltime savings as their operational goal, the HOV restrictions should be imposed only during those time periods when these savings can be achieved.
- The entire project should be opened at one time (at least by direction).
- Priority sections should be particularly well maintained.
- Enforcement should be supported by extensive public education and publicity of the seriousness of the HOV restrictions.
- Aggressive enforcement should begin immediately to instill a degree of respect of the HOV restrictions.
- A readily accessible refuge area (full shoulders) should be provided for stationary observation and apprehension. If this is not possible, serious consideration should be given to extensive selective, special, or instrumented enforcement tactics.

CHARACTERISTICS OF SUCCESSFUL HOV PROJECTS

A list of the characteristics and strategies associated with successful HOV lane projects includes:

- Physically separated lanes representing added capacity to existing roadway
- Exclusive entrance and exit ramps
- The addition an HOV lane where added capacity is needed
- Provision of benefits to HOV lane users without losses to regular lane traffic
- Use of an HOV lane where significant time savings for users can be obtained; e.g., 1 minute per mile or a cumulative total of 10 minutes or more
- Except where used as a bottleneck bypass, the greater length of an HOV lane tends to produce greater cumulative time saving benefits
- Provision of a breakdown lane serving the HOV lane for safety and enforcement
- Refuge cutouts for breakdowns, when a shoulder is used to add HOV lane capacity
- Limitation of HOV operating times to the peak hours to avoid unused capacity and adverse public reaction
- Designation of an HOV lane for exclusive use by buses (busway) when bus utilization is sufficiently high to absorb most HOV lane capacity
- Limitation of carpool definition to two, three, or four persons based on existing demand and conditions of congestion, likelihood of HOV lane capacity being reached, and efficient utilization of HOV lane capacity
- Definition of HOV lane users in a manner to ensure that a high percentage of vehicles qualify at the tiem of implementation
- Generation of laws and regulations that support enforcement and implementation of HOV lanes
- Adequate, frequent, consistent, and visible enforcement
- An informed and sympathetic court system
- Involvement of all agencies likely to impact or be impacted by HOV lane operation in the planning and implementation program
- Extensive marketing and public information program prior to and during HOV lane implementation
- Combination with other features for synergistic effect: i.e., park-and-ride parking lots, added express bus service, toll reductions or elimination for HOV's, exclusive busways downtown, metered ramps with HOV bypass lanes, carpool-matching and incentive programs, and flexible working hours

REFERENCES

1. D. Baugh & Associates, Inc. "Freeway High-Occupancy Vehicle Lanes and Ramp Metering Evaluation Study." Prepared for the U.S. Department of Transportation. March 1979.

2. Beiswenger, Hock & Associates. "Enforcement Requirements for High-Occupancy Vehicle Facilities."

3. JHK & Associates, Wagner McGee Associates, and the Traffic Institute at Northwestern University. FHWA training course in High Occupancy Vehicle Facility Development, Operation and Enforcement.

CHAPTER 10. DETECTION OF NON-RECURRING CONGESTION (INCIDENTS)

The ability to detect incidents quickly is undoubtedly one of the most powerful techniques for reducing delay and improving overall freeway operation.

Various surveillance systems are used for incident detection. Some systems convey information concerning the type of response required to remedy the problem that has been detected. In most cases, however, the detection process is an independent operation and some form of follow-up is required to ascertain the nature and extent of the incident and the type of response required. The most common surveillance systems use:

- Motorist call systems
- Citizens band radio
- Police and service patrols
- Aerial surveillance
- Closed-circuit television
- Detector-based surveillance

MOTORIST CALL SYSTEMS

One of the earliest incident detection systems used motorist callboxes or emergency telephones. Motorists experiencing, or witnessing, an incident used the nearest callbox or telephone to inform the operating agency of the nature of the incident. Telephones are generally preferred because voice communication gives the motorist an opportunity to explain exactly what services are required. However, the callbox with coded message buttons is less costly than a telephone requiring voice transmission.

The major advantage of a motorist call system is that it is an efficient system for signaling a motorist's need for service. A major disadvantage is the delay associated with the motorist's determining that an incident has occurred, determining that the proper action involves using the callbox, locating the nearest callbox, and then proceeding safely to the callbox to inform the operating agency. This delay can be quite significant. Another major disadvantage is the large number of "gone-on-arrival" calls that are generated (i.e., the motorist remedies his problem through some other means and when the servicing agency arrives the disabled vehicle is no longer there).

CITIZENS BAND RADIO

Another way freeway incidents can be detected is through the use of citizens band (CB) radio. Drivers of vehicles equipped with a CB radio observed and reported incidents to a central monitoring center which, in turn, transmits the information to the appropriate agency for dispatch of the required assistance. The key elements of this system are motorists who are knowledgeable about the system and willing to report the incidents they observe. Signs informing the motorists which CB channels are monitored are necessary. Because of this volunteer aspect, the detection capability of the system is always a function of the number of motorists on the freeway who have the necessary CB equipment and are willing to provide their cooperation. With the growing number of CB-equipped vehicles, this type of system has considerable potential. This system has been working well in Detroit where remote CB units are used to monitor calls. These remote units are called up over telephone lines when an incident is detected and the broadcast is then monitored to provide additional information on the nature of the incident. In Minneapolis, two remote CB units are continuously monitored and initial results show that one in six incidents were detected over the CB radio before they were detected by the automatic incident detection system.

POLICE AND SERVICE PATROLS

Various patrol systems have been used on urban freeways to provide incident detection. The most common is the use of police patrols that circulate in the traffic stream and have as their primary objective the identification of incidents, determination of the nature and extent of the incident, and dispatch of the type of response needed. The major advantage of police patrols is that detection and dispatch of response is one function. The major disadvantage is the large number of patrols required to effectively cover a freeway system and the high costs involved.

Another system used to provide incident detection is the service patrol. This system involves the use of light-duty service vehicles and, similar to the police patrols, provides for detection of incidents. It also provides minor services such as fuel, oil, water, and minor mechanical repairs. As with the police patrols, this system is relatively expensive because of the large number of patrols required. The patrol vehicles in regular use in Chicago have assisted over 1 million motorists in the past 20 years.

Since surveillance techniques are complementary, most operational surveillance systems utilize more than one concept to achieve a cost-effective surveillance and control system.

AERIAL SURVEILLANCE

This type of surveillance is primarily used by police and commercial radio stations to get a general overview of traffic in a particular area or corridor. Through the use of light planes or helicopters, they observe where the bottlenecks are occurring and determine whether they have been caused by incidents. Advisories of this information are then broadcast to motorists, and assistance is dispatched to the scene of the incident. Due to the expense of this type of surveillance, usually a wide geographical area must be covered. Consequently, there often is considerable time delay in identification and removal of incidents.

In general, it has never been conclusively demonstrated that aerial surveillance is a cost-effective technique for incident detection. The equipment and the labor requirements of the system are expensive and its effectiveness is sometimes limited by weather conditions.

CLOSED-CIRCUIT TELEVISION

Incident detection by closed-circuit television is performed using operators in a central control room who monitor traffic conditions using televison cameras placed at critical locations on the freeway. When an incident occurs, the operator determines the nature of the incident and the type of response likely to be required. In general application, closed-circuit television is limited to those locations where delay-causing incidents are a chronic problem and fast response is essential. In this use, the television normally serves to confirm electronic surveillance where incidents are detected automatically and an alarm is used to alert the operators. It also provides information on the nature of the incident.

The major advantage of closed-circuit televison is that it provides a full view of a section of freeway. The major disadvantages are:

- It is expensive to install and maintain. However, recent technological developments may alleviate this problem as the costs of video cameras fall and communication systems become more reliable.
- It is often difficult and expensive to obtain good pictures under adverse weather conditions and after dark. However, new technology has resulted in provision of much better pictures under such conditions.
- Monitoring of the TV screen is a tedious task. Without an automatic alarm in a detector-based surveillance system, operators tend to lose interest and consequently fail to notice incidents immediately.
- Continuous monitoring of TV screens by qualified operators is expensive.

The value of closed-circuit television is primarily to confirm and provide information about the nature of incidents and other factors causing congestion in known areas of recurring incidents. It has shown to be very useful in this regard, and there is a general concensus that television surveillance is a necessary element of an urban freeway system, especially if it is used on a selective—not systemwide—basis. In many freeway management systems, television is essential if the surveillance system is to attain credibility with police and others charged with responding to surveillance alarms.

DETECTOR-BASED SURVEILLANCE

Incident detection by electronic surveillance is real-time monitoring of traffic data through use of detectors installed at critical locations along the freeway. When a delay-causing incident occurs, the capacity of the freeway is reduced at the point of occurrence. If capacity is reduced to some figure less than the existing demand, the traffic flow upstream of the incident will be affected. If the changes in traffic flow are greater than some predetermined value, it is likely that an incident has occurred. In this manner, incidents are detected by logically evaluating the variations in flow characteristics. (Some work has been conducted using speed as the determining variable, but most electronic surveillance systems use occupancy for incident detection.)

In the Los Angeles Freeway Surveillance and Control Project, changes in the percentage of time that vehicles spend above a detector (lane occupancy) are used to sense congestion and indicate that an incident has occurred. A computer calculates the difference in occupancy between adjacent detector stations (in this case, spaced at 1/2-mile intervals). At the end of each sampling period and when the relative percent change between the present occupancy and that of the preceding sample for the downstream detectors exceeds a certain value, an alert is signaled automatically by the computer. Additional information on traffic conditions immediately upstream of the incident is then obtained, and decisions are made as to what response is needed.

The main advantage of detector-based surveillance is that it provides a continuous network-wide monitoring capability at a relatively low operating cost. It can also be used for many other tasks such as establishing metering rates for traffic-responsive ramp-metering systems. Its main disadvantage is that the nature of the incident cannot be readily identified, and some follow-up surveillance is often required to determine what response is needed. The use of incident detection algorithms in electronic surveillance is discussed in the next chapter.

CHAPTER 11. INCIDENT DETECTION ALGORITHMS

INTRODUCTION

Incident detection algorithms are the logic used to evaluate the information obtained from electronic surveillance. As discussed in chapter 10, one of the primary objectives of electronic surveillance is the detection of non-recurring incidents. Incidents of this type, such as accidents, stalled vehicles, and spilled cargoes, occur randomly and are unpredictable.

Although incidents are statistically rare events, occurring once every 20,000 to 30,000 vehicle-miles on high-volume urban freeways, they occur with such frequency that they must be taken into account by surveillance and control systems(1). West(2) indicates that the non-recurring freeway congestion due to incidents is responsible for as much motorist delay in the urban area as is the recurring congestion due to geometric bottlenecks.

Evaluating Incident Detection Algorithms(3)

The measures of effectiveness generally used in evaluating detection algorithms are detection rate, false alarm rate, and mean time to detect.

- Detection rate is the percentage of incidents detected out of all capacity-reducing incidents that occur during a specified time period.
- False alarm rate is the percentage of false incident messages out of the total messages during a specified time period.
- Mean time to detect is the mean delay between the actual occurrence of an incident and the time when it is detected for all incidents during a certain time period.

When choosing an algorithm there is normally a trade-off between these measures of effectiveness. In most cases, to achieve a high detection rate the engineer must settle for a higher false alarm rate and vice versa (for a low false alarm rate a lower percentage of the total incidents will be detected). To lower response time, the engineer must accept a higher false alarm rate and the cost of additional computations.

Increasing the frequency of algorithm execution will not produce great improvements in response time. It has been found(3) that the reduction in response time when tests are performed every x seconds, rather than every 60 seconds, is given by (60-x)/2 seconds. Therefore, executing an algorithm every 20 seconds will lower the response time by an average of 20 seconds. Due to the marginal benefits obtained from reducing the execution interval, a 60-second interval is generally recommended.

For purposes of selecting an algorithm, the false alarm rate is likely to be the controlling measure. The engineer should select an algorithm with a false alarm rate not greater than that which can be tolerated, and accept the corresponding detection rate and mean time to detect.

The first step is to estimate the tolerable false alarm rate. To do this, these elements must be considered: the number of sections of roadway under surveillance, duration of the peak period and the availability of other resources for identifying and verifying incidents, particularly TV surveillance. The duration of the peak period is relevant because incidents are more likely during this period when they have their greatest disruptive impact.

A second factor to consider when selecting an algorithm is the potential presence of compression waves in the traffic stream. Compression waves occur in heavy, congested traffic and are associated with severe slow-down, speed-up vehicle speed cycles. A compression wave is characterized by a sequence of large values of occupancy, greater than 30 percent, which move in the upstream direction as time progresses. Table 11.1 presents 1-minute occupancy data for a section of the San Diego freeway in Los Angeles with an example of a compression wave. If compression waves are a common problem, this fact should be accommodated when selecting an algorithm.

Algorithm Logic

Incident detection algorithms are normally based on the general structure of the binary decision tree.

A binary decision tree consists of a set of nodes connected by links to form a tree-like structure. A node is either a decision node or a terminal node. A decision node consists of a comparison of a feature to a threshold and a specification to two successor nodes—one which is next to be examined if the comparison is true, and the second, if the comparison is false. The elementary building block of a binary decision tree is illustrated in figure 11.1. Each successor node may be a decision node or a terminal node.

The features used in the decision nodes of the binary decision tree are usually based on the variable occupancy and volume, which are defined as:

- OCC(i,0,t) – The 1-minute occupancy, expressed as a percentage, at station i, averaged across all lanes
- OCC(i,j,t) – The 1-minute occupancy, expressed as a percentage, at station i, lane j, for $j \geq 1$
- VOL(i,0,t) – The 1-minute count, expressed as vehicles per lane per hour, at station i, and averaged across all lanes
- VOL(i,j,t) – The 1-minute count, expressed as vehicles per lane per hour, at station i, lane j, $j \geq 1$

For example, using figure 11.2 to illustrate, OCC(1,2,t) is the occupancy in lane 2 of station 1; OCC(2,0,t) is occupancy averaged across all lanes at station 2. The convention used here will be that station i is the upstream station and station i+1 is the downstream station. The number of lanes at station i is denoted by N(i).

Of all the algorithms in use, the most widely used and recognized was developed by the California Department of Transportation and first implemented on 42 miles of the San Diego, Santa Monica, and Harbor Freeways. The California algorithm consists of three comparisons to preset threshold values that must all be exceeded before an incident is signaled. Where T_1, T_2, T_3 are predefined, station-

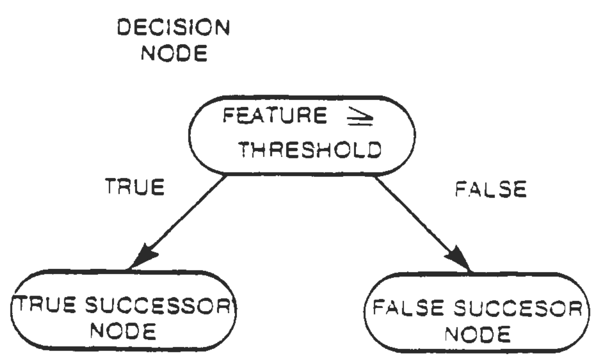


Figure 11.1. Elementary Building Block of a Binary Decision Tree

TABLE 11.1. ONE-MINUTE OCCUPANCY DATA EXHIBITING COMPRESSION WAVES(3)

Direction of Travel →

Time	Station 32	31	30	29	28	27	26
710	15	20	20	18	22	26	22
711	13	21	19	18	22	28	26
712	16	19	20	19	21	32	26
713	13	18	16	18	30	25	25
714	14	22	17	18	25	23	24
715	14	20	20	26	44	29	26
716	14	18	18	25	34	26	24
717	13	21	19	36	26	21	25
718	14	24	21	48	29	25	21
719	16	26	32	28	31	26	25
720	21	24	47	19	26	39	23
721	14	26	32	27	26	21	22
722	14	52	32	22	29	19	23
723	14	27	23	20	$\boxed{50}$	18	24
724	13	26	21	21	30	22	26
725	24	21	22	$\boxed{62}$	23	26	24
726	39	20	23	38	23	28	23
727	23	21	$\boxed{65}$	29	22	30	23
728	26	24	43	28	23	23	25
729	31	26	26	29	22	30	23
730	30	$\boxed{60}$	22	35	22	24	23
731	31	41	21	30	17	26	24
732	37	29	27	26	23	18	26
733	$\boxed{50}$	26	35	22	37	22	24
734	53	22	31	21	29	26	26
735	48	21	32	21	25	22	23
736	29	28	33	39	21	24	23
737	37	33	28	26	22	30	27
738	38	29	44	21	20	23	24
739	40	25	38	21	21	20	27
740	53	23	43	19	30	23	24
741	37	47	44	22	36	26	23
742	41	30	42	23	38	28	26
743	38	26	38	21	31	22	25
744	56	24	29	33	29	23	22
745	64	25	24	38	27	27	24

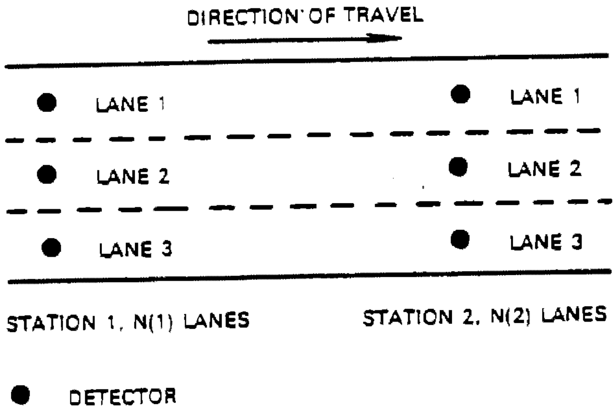


Figure 11.2. Lane Sensor Notation

specific thresholds, the three comparisons are:

1. $OCC(i,0,t) - OCC(i+1,0,t) \quad T_1$

2. $\dfrac{OCC(i,0,t) - OCC(i+1,0,t)}{OCC(i,0,t)} \quad T_2$

3. $\dfrac{OCC(i+1,0,t-2) - OCC(i+1,0,t)}{OCC(i+1,0,t-2)} \quad T_3$

Figure 11.3 shows the California algorithm as a binary decision tree.

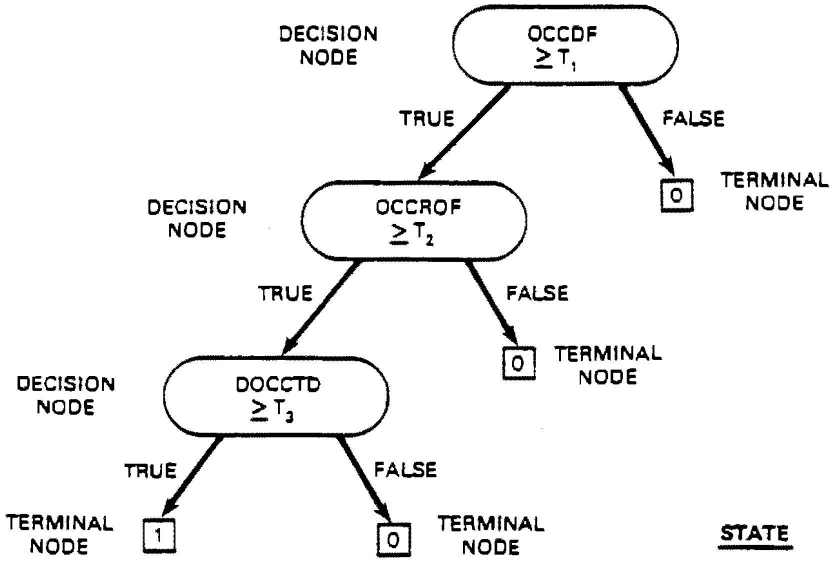
During the first 7 months of evaluation of the Los Angeles Area Freeway Surveillance and Control Project, 2,700 congestion-causing events were verified on the 42-mile loop. Slightly more than 90 percent were detected by the California algorithm. The false alarm rate (percent of false incident messages of total incident messages) was about 32 percent for all periods of the day. It was approximately 38 percent during peaks and 19 percent during off-peak periods. Considering all verified incidents, more than 80 percent of all incidents were detected in under 4 minutes(4).

For any algorithm that the engineer uses, the set of thresholds can be varied to produce differing performance. The engineer will have to decide what is an acceptable false alarm rate, set the thresholds that will produce that false alarm rate and accept the corresponding detection rate. The false alarm rate that will be acceptable will be highly dependent on the availability of other resources for identifying and verifying incidents, particularly TV surveillance. If there is TV surveillance, a much higher false alarm rate will be tolerable since the incident can be verified by simply looking at the appropriate TV monitor. If the police or other personnel must be sent to the site to verify an incident, the cost associated with responding to a high number of false incidents will dictate a lower false alarm rate.

Payne and Knobel evaluated a modified version of the California algorithm on Los Angeles and Minneapolis freeway surveillance data. The modified version involved a test for termination of an incident and is shown in figure 11.4. The results of the evaluation for seven sets of threshold values are shown in table 11.2. The results in table 11.2 illustrate the tradeoffs discussed earlier with regard to detection rate, false alarm rate, and mean time to detect.

Payne and Knobel tested nine other algorithms besides the modified California model. The other algorithms examined ranged from simple modifications of the California algorithm to complex algorithms with test for compression waves and persistence. Of the simple algorithms evaluated by Payne and Knobel the one shown in figure 11.5, called algorithm 7, was found to be the best. The definition of the features used in algorithm 7 and the other algorithms evaluated by Payne and Knobel are given in table 11.3

Algorithm 7 is similar to the California algorithm except it uses DOCC instead of DOCCTD as one of its tests. It also re-



STATE	DESIGNATES
0	INCIDENT-FREE CONDITION
1	INCIDENT CONDITION

OCC = MINUTE AVERAGE OCCUPANCY (EXPRESSED AS A PERCENTAGE)

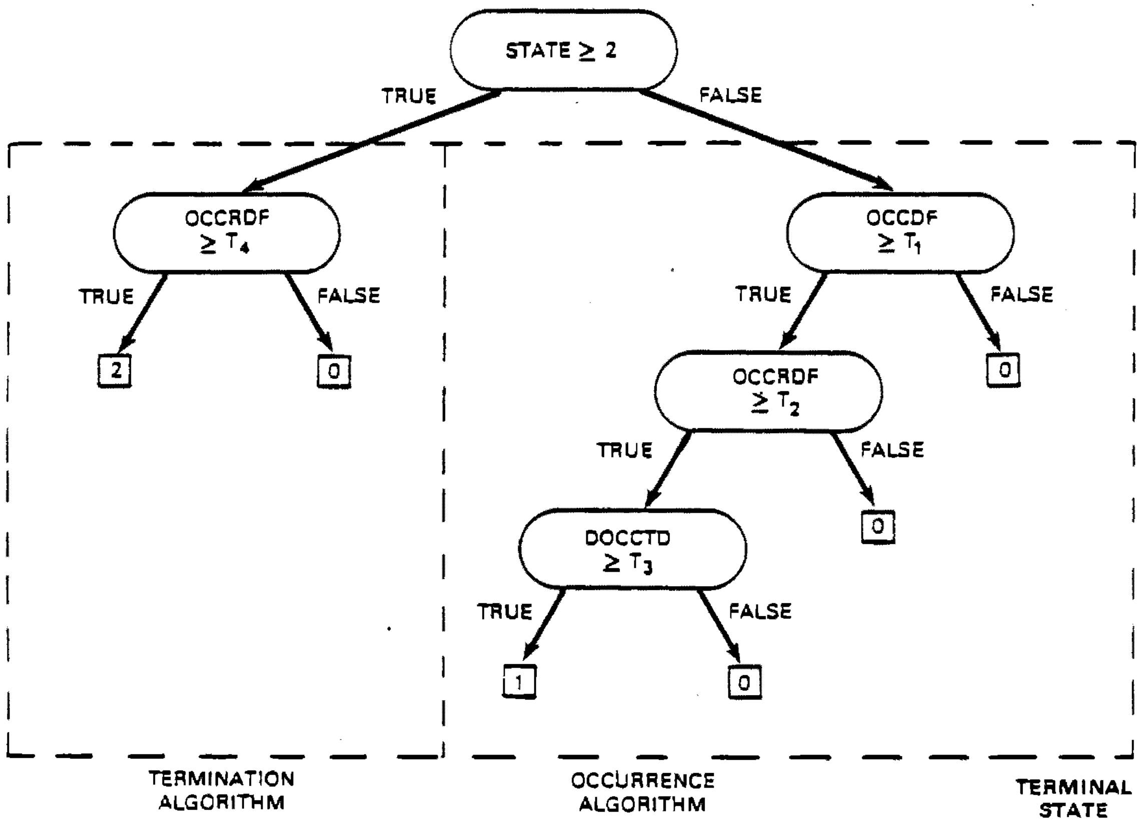
OCCDF = OCC (UPSTREAM) - OCC (DOWNSTREAM)

$$\text{OCCRDF} = \frac{\text{OCCDF}}{\text{OCC (UPSTREAM)}}$$

DOCCTD = [OCC (DOWNSTREAM, TWO MINUTES AGO) -OCC (DOWNSTREAM NOW)]/ OCC (DOWNSTREAM, TWO MINUTES AGO)

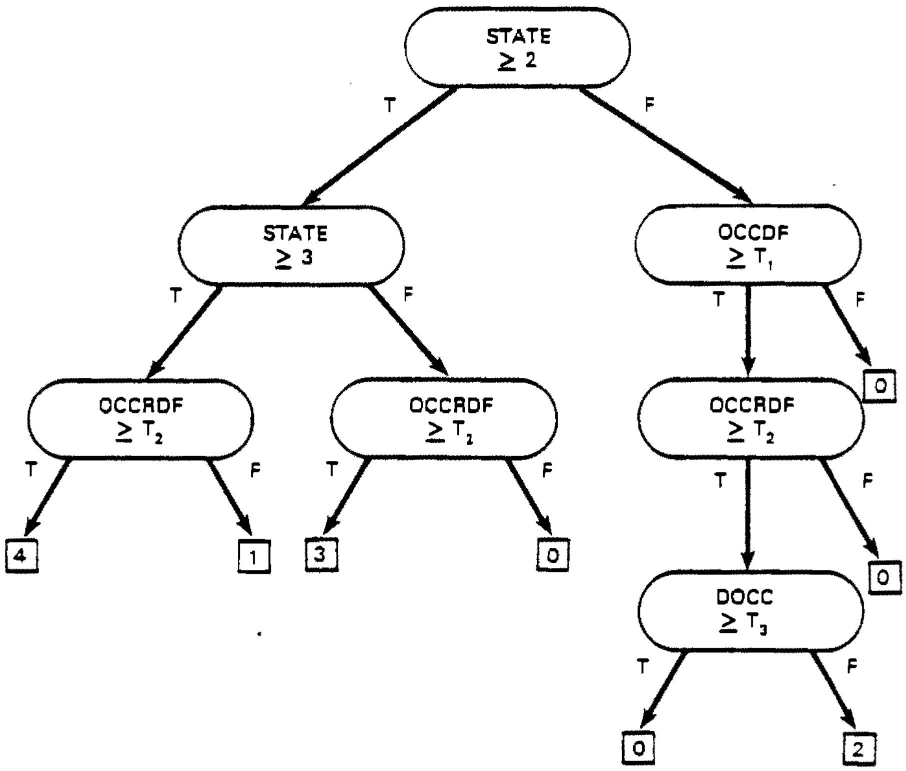
T_1, T_2, T_3 = THRESHOLDS

Figure 11.3. The California Algorithm



TERMINAL STATE VALUE	DESIGNATES
0	INCIDENT-FREE
1	INCIDENT OCCURRED
2	INCIDENT CONTINUING

Figure 11.4. The California Algorithm with a Test for Termination (3)



STATE	DESIGNATES
0	INCIDENT-FREE
1	INCIDENT TERMINATED
2	TENTATIVE INCIDENT
3	INCIDENT OCCURRED
4	INCIDENT CONTINUING

Figure 11.5. Decision Tree for Algorithm 7, 60-Second Updating

TABLE 11.2. PERFORMANCE OF MODIFIED CALIFORNIA ALGORITHM ON LOS ANGELES DATA BASE(3)

Threshold Set	Thresholds T_1	T_2	T_3	Detection Rate (Percent)	False Alarm Rate (Percent)	Mean Time to Detect (Minutes)
1	5.3	.308	.061	82	1.294	0.85
2	5.8	.340	.112	71	.901	1.99
3	7.7	.498	.049	61	.309	3.33
4	5.0	.563	.013	51	.222	5.09
5	9.6	.617	.075	41	.070	5.18
6	15.3	.637	.245	31	.015	6.00
7	13.0	.710	.192	20	.004	7.77

TABLE 11.3. DEFINITIONS OF SIMPLE FEATURES

Feature	Description	Definition
DOCC(i,t)	Downstream occupancy	OCC(i+1,0,t)
OCCDF(i,t)	Spatial difference in occupancies	OCC(i,0,t) - OCC(i+1,0,t)
OCCRDF(i,t)	Relative spatial difference in occupancies	$\frac{\text{OCCDF}(i,t)}{\text{OCC}(i,0,t)}$
DOCCTD(i,t)	Relative temporal difference in downstream occupancy	$\frac{\text{OCC}(i+1,0,t-2)-\text{OCC}(i+1,0,t)}{\text{OCC}(i+1,0,t-2)}$

quires persistence of the discontinuity for 1 minute. The results of algorithm 7's performance on Los Angeles data is given in table 11.4.

For freeways where compression waves are a problem, algorithm 8 (also tested by Payne and Knobel) was found to be most effective. The structure of algorithm 8 is shown in figure 11.6 and the results of its evaluation are given in table 11.5. The relative temporal difference in downstream occupancy, DOCCTD, is the primary logic used to suppress a compression wave.

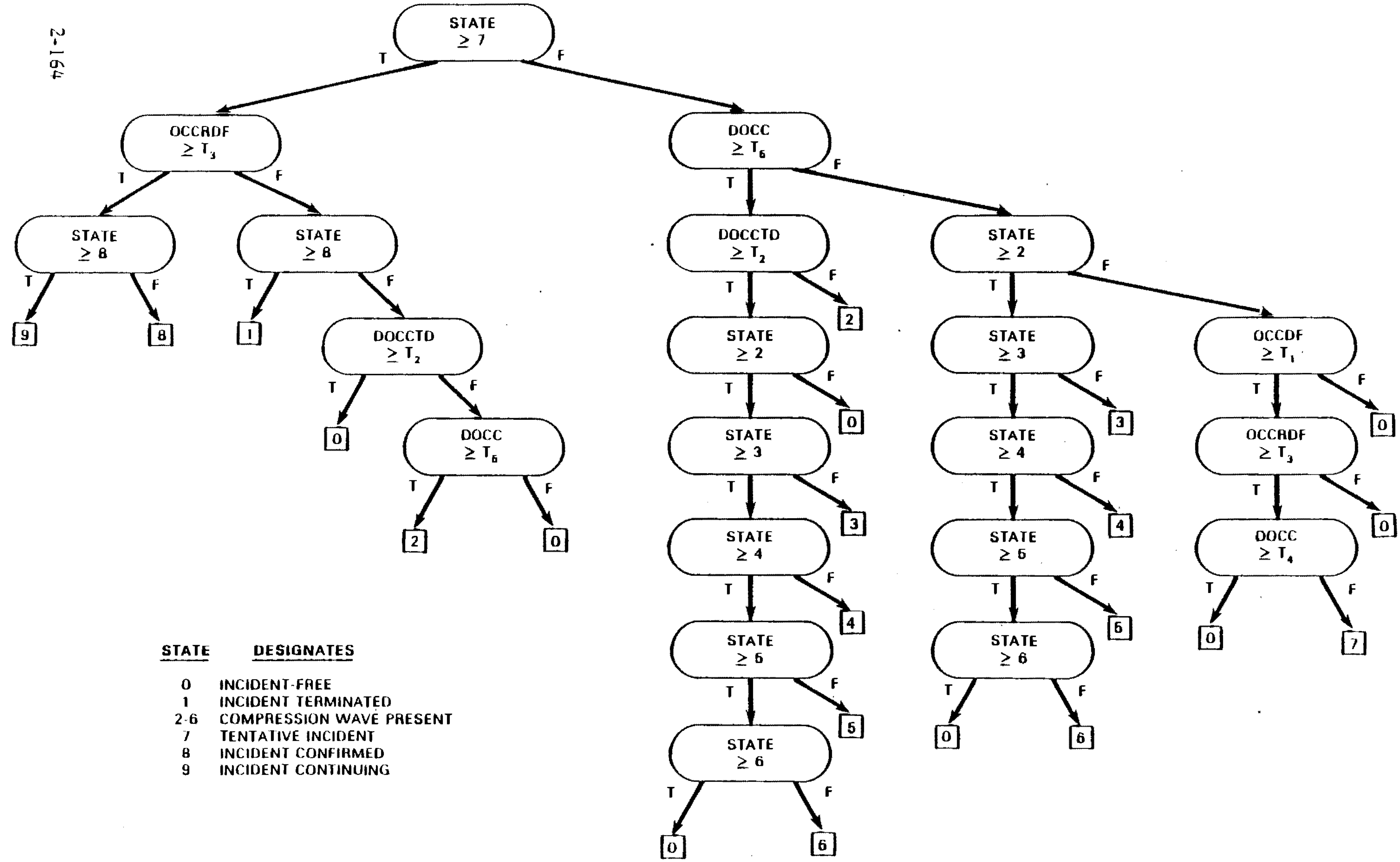
Cook and Cleveland(1) reported the results of their evaluation of a double exponential smoothing model. They found there was a definite advantage in considering more than the last one or two occupancy data observations when trying to determine the current traffic flow trends. In this method, a tracking signal is used to consider the statistical magnitude of data noise before a decision is made that an

TABLE 11.4. PERFORMANCE OF ALGORITHM 7
ON LOS ANGELES DATA BASE(3)

Threshold Set	Thresholds T_1	T_2	T_3	Detection Rate (Percent)	False Alarm Rate (Percent)	Mean Time to Detect (Minutes)
1	8	.31	17	59	.134	3.25
2	13	.36	17	51	.050	4.31
3	13	.39	13	37	.017	6.17
4	22	.30	14	31	.006	5.84
5	27	.32	13	20	.004	7.73

TABLE 11.5. PERFORMANCE OF ALGORITHM 8
ON LOS ANGELES DATA BASE

Threshold Set	Thresholds T_1	T_2	T_3	T_4	T_5	Detection Rate (Percent)	False Alarm Rate (Percent)	Mean Time to Detect (Minutes)
1	10	-.44	.31	29	30	61	.177	3.88
2	13	-.30	.31	16	30	51	.038	4.79
3	18	-.31	.36	19	30	41	.024	5.63
4	24	-.39	.58	13	30	20	.003	8.83



STATE	DESIGNATES
0	INCIDENT-FREE
1	INCIDENT TERMINATED
2-6	COMPRESSION WAVE PRESENT
7	TENTATIVE INCIDENT
8	INCIDENT CONFIRMED
9	INCIDENT CONTINUING

Figure 11.6. Algorithm 8, 60-Second Updating, 5-Minute Suppression

incident has occurred. The traffic parameter occupancy is smoothed as follows:

$$S1(t) = \alpha occ(t) + (1-\alpha) S1(t-1)$$

$$S2(t) = \alpha S1(t) + (1-\alpha) S2(t-1)$$

Where:

$S1(t)$	=	Single exponential smoothing formula for the time interval t
$S2(t)$	=	Double exponential smoothing formula for the time interval t
$occ(t)$	=	Occupancy for time interval t
α	=	Smoothing parameter

With the results of these two formulas, a predicted linear forecast of occupancy for the time interval t+1 can be made. For each time interval t, a tracking signal is computed for the purpose of determining if the observed value of occupancy could have occurred by chance or if it reflects the presence of a roadway incident. Cook and Cleveland concluded that of all incident detection models they evaluated, the exponentially smoothed occupancy model was the most effective(5).

Dudek and Meanes(6) developed and evaluated an incident detection algorithm using the standard normal deviation (SND) of the control variable (energy or lane occupancy). A high rate of change in the control variable will reflect on the incident situation as distinguished from a normal demand-capacity problem caused by geometrics. The SND is a standardized measure of the deviation from the mean in units of the standard deviation and is expressed by the following relationship:

$$SND = \frac{x - \bar{x}}{s}$$

Where:

x	=	value of control variable at time t
$\bar{x}$	=	mean of control variable over previous n sampling periods
s	=	standard deviation of control variable over previous n sampling periods

The value of SND will reflect the degree to which the control variable has changed in relationship to the average trends measured during previous intervals. A large SND value would reflect a major change in operating conditions on the freeway.

The SND model was incorporated with a stoppage-wave detection algorithm previously developed for operating safety warning devices on the Gulf Freeway in Houston(7,8). The model was evaluated using two strategies: strategy A requires one SND value to be critical while strategy B requires two successive SND values to be critical. The evaluation also used two control variables, occupancy and energy, and 3- and 5-minute time bases. The structure of strategies A and B is shown in figure 11.7. The time base is the period of time over which the mean and standard deviation is calculated. Cumulative distributions of the percent of incidents detected and the percent of false alarms using strategy B with lane occupancy as the control variable with a 5-minute time base are shown in figure 11.8. The results of the different strategies, control variables, and time bases are given in table 11.6

Several other algorithms(9,10,11) have been developed and evaluated while others are still being tested(12). At this time, no one algorithm has been found to be superior to all others. The main problem with incident detection algorithms is still the high number of false alarms.

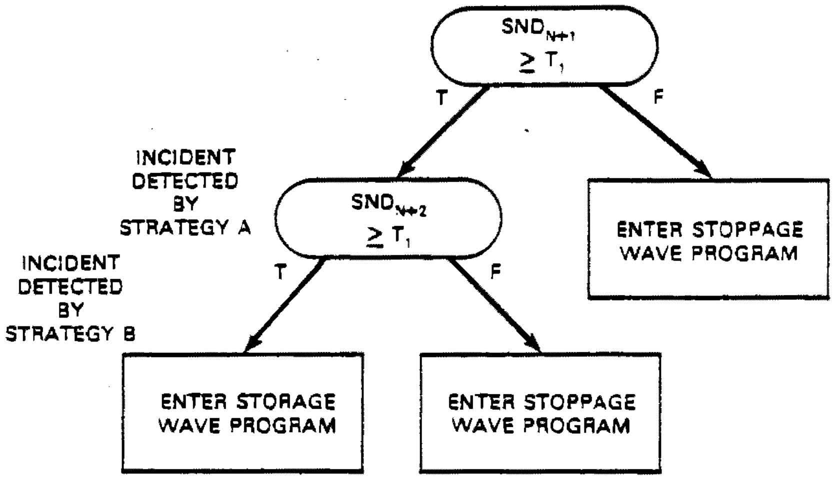


Figure 11.7. Strategies A and B for SND Incident Detection Model (6)

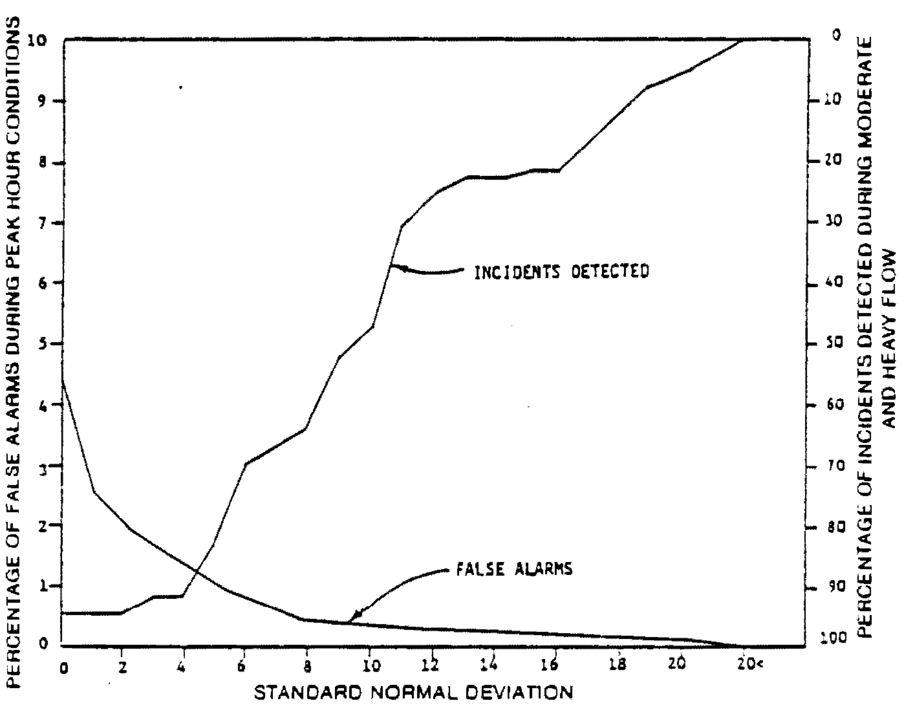


Figure 11.8. Performance Curves Using Lane Occupancy, Strategy B and 5-Minute Time Base

TABLE 11.6. EFFECTIVENESS OF SND DETECTION STRATEGIES

Strategy	Variable	Time Base (Minutes)	Critical SND Values	Average Computer Response Time (Minutes)	Standard Deviation Computer Response Time	Incidents Detected[a] (Percent)	False Alarms[b] (Percent)
A	Occupancy	5	6	0.5	1.1	86	1.7
	Occupancy	3	6	0.7	1.9	86	2.0
	Energy	5	-4	0.8	2.6	86	2.4
	Energy	3	-3	0.3	0.7	86	2.5
B	Occupancy	5	4	1.1	0.6	92	1.3
	Occupancy	3	4	1.1	1.5	89	1.4
	Energy	5	-3	1.1	0.5	83	1.4
	Energy	3	-3	1.1	0.5	83	1.4

a. During moderate and heavy flow conditions

b. During peak periods

REFERENCES

1. Cooke, A. R., and Cleveland, D. E. "Detection of Freeway Capacity-Reducing Incidents by Traffic-Stream Measurements." Transportation Research Record No. 495, 1974, pp. 1-11.

2. West, J. "Proposed Real-Time Surveillance and Control System for Los Angeles." Paper presented to HRB Committee on Freeway Operations. August 1969.

3. Payne, H. J., and Knobel, H. C. "Development and Testing of Incident Detection Algorithms, Volume 3: User Guidelines," Technology Service Corporation. April 1974.

4. Green, Richard H. "The 42-Mile Freeway Surveillance Loop." Freeway Operation Department, District 7, California Division of Highways. 1973.

5. Tignor, S. C., and Payne, H. J. "Improved Freeway Incident Detection Algorithms," Public Roads, Vol. 41, No. 1, June 1977, pp. 32-40

6. Dudek, C. ., and Messer, C. J. "Incident Detection and Urban Freeways." Transportation Research Record No. 495, 1974, pp. 12-24.

7. Dudek, C. L, Messer, C. J., and Friebele, J. D. "Investigation of Lane Occupancy as a Control Variable for a Safety Warning System for Urban Freeways." Texas Transportation Institute, Reserach Report No. 165-6. March 1973.

8. Dudek, C. L., and Messer, C. J. "Detecting Stoppage Waves for Freeway Control." Highway Research Record No. 469, 1973, pp. 1-15.

9. Courage, K. G., and Levin, M. "A Freeway Corridor Surveillance, Information and Control System." Research Report No. 488-8. Texas Transportation Institute. December 1968.

10. Dudek, C. L., Weaver, G. D., Ritch, G. P., and Messer, C. J. "Detecting Freeway Incidents Under Low-Volume Conditions." Transportation Research Record No. 533, 1975, pp. 34-47.

11. Sakasita, M., and May, A. D. "Development and Evaluation of Incident-Detection Algorithms for Electronic-Detector Systems in Freeways." Transportation Research Record No. 533, 1975, pp. 48-63.

12. Collins, J. F., Hopkins, C. M., and Martin, J. A. "Automatic Incident Detection — TRRL Algorithms HIOCC and PATRES." Transport and Road Research Laboratory Supplementary Report No. 526, 1979.

CHAPTER 12. PLANNING FOR INCIDENT RESPONSE

INTRODUCTION

This chapter describes the various ways in which incident response can be organized and performed, and some additional administrative issues that should be considered when planning for incident response. These options are not mutually exclusive, and it is likely that several would be incorporated into the freeway management system. The final section of the chapter contains an explanation of the evaluation process for selecting between alternative groups of options that may be implemented.

INCIDENT MANAGEMENT

The national cost of freeway incidents exceeds $100 million dollars yearly. This cost can be measured in terms of wasted gasoline, loss of productive time, and air pollution. Due to the random nature of spills, breakdowns, and accidents, nonrecurring problems cannot be remedied by using of the same solutions suitable for recurring problems. A cost effective approach to freeway incident management is particularly suitable in times of economic restraint, and can be used as an interim solution to the problem of recurring congestion.

The type and extent of existing freeway incident management systems throughout the country depend greatly on the public agencies or organizations operating them. Two factors exert a strong influence on the nature of an agency's involvement: its established role and expertise, and its current perception of the freeway incident management problem. In considering either the establishment of a new system or the improvement of an existing one, engineers will be influenced by these factors. The five types of operating agencies or groups that are usually involved in freeway incident management are:

- Police agencies
- Departments of transportation
- Point facility (e.g., bridges and tunnels) authorities
- Citizen groups
- Freeway incident management teams

In most cities, the state or local police agency has operational responsibility for the freeway system. This duty evolves quite naturally from traditional police concerns of traffic law enforcement, traffic control, and accident investigation. Some police departments view their role as control over the vehicles in the freeway environment, in contrast to the highway department's responsibility for the stationary elements. Given this high degree of involvement in freeway operations, many police agencies regard incident management as a logical extension of their existing responsibilities.

Police agencies tend to implement incident management techniques that build on their basic patrol functions. A meaningful commitment requires the assumption of a service role. This may not be practical for some police agencies, due to limited experience or inadequate resources. Under such circumstances, incident management responsibility may fall on the next most logical public agency, a highway department or department of transportation (DOT).

The traditional construction and maintenance responsibilities of highway departments and DOT's have necessitated the acquisition of many of the vehicles and much of the equipment used for the cleanup and removal of freeway incidents. Furthermore, some DOT's have acquired an additional incident management role as a result of operating electronic surveillance and control systems. These activities provide a sound base upon which to build a more comprehensive incident management system under DOT control. Many of the potentially new incident management techniques, however, involve responsibilities that are usually taken by police agencies.

A DOT-sponsored incident management system can only be seriously considered with the expressed approval and cooperation of the local police agency. It can operate effectively under two different organizational configurations: as the joint responsibility of the DOT and the police department, with each agency still performing its traditional role but having closer coordination with the other; or as the sole responsibility of the DOT, with the agency assuming what were previously police incident management responsibilities (such as providing motorist-aid services or monitoring CB radio reports).

For bottleneck or point facilities such as bridges and tunnels, the operating authority is often the initiator of an incident management system. These facilities present special problems, in that shoulders or emergency service lanes are usually nonexistent and diversion opportunities limited. Consequently, measures to reduce delay and congestion are of great importance to the authority, which, because of its operational experience, is the most logical group to oversee implementation. Furthermore, police agencies and highway departments generally play only supporting roles with respect to incident management on point facilities. Point facility incident management systems tend to emphasize their surveillance measures, since detection over a short stretch of highway can be virtually instantaneous without being prohibitively expensive. This fact explains the low-cost classification of options using closed-circuit television and loop detectors for point facilities, even though their installation throughout an entire freeway system would require a large investment.

A citizen group may take the lead in implementing incident management options, particularly when official activities are perceived as inadequate. Because of its essentially volunteer nature and limited resources, this option tends to be labor-intensive and often consists of methods for improving the speed and quality of the incident-related information transmitted to public agencies and the media. Citizen incident management involvement frequently develops out of existing organizations such as citizen band (CB) radio clubs and civic associations.

The freeway incident management team is considered to be the most effective organization due to the consolidation of all response and surveillance responsibilities into one administrative unit. This allows for the complete management of incidents by a single multidisciplinary team composed of police, highway, and other public agency personnel, with possible assistance by citizen volunteers. Such an authority is usually not available, so a team system is likely to require an intense, well planned organization and administrative effort at its inception. The elimination of coordination and communication problems that reduce the effectiveness of other systems, however, makes the extra effort worthwhile. An additional advantage of the team system is that all incident management options are available to it, since it represents a combination of the other four systems. The advantages of a team system and the processes for interagency cooperation are also discussed in chapter 16 of this volume.

Figure 12.1 illustrates the options available for incident management. The figure shows which agencies generally have the operational responsibilities for each of the options. These divisions are by no means rigid, as in specific locations one agency often takes on the responsibility for many of the options listed. In the case of point facilities at bridges or tunnels, the operating agency can be considered to perform functions that are similar to the functions of a highway department.

ADDITIONAL INCIDENT DETECTION, MANAGEMENT, AND ADMINISTRATIVE OPTIONS

Following a positive response to whether a breakdown service may be warranted, an engineer should investigate the various options(1) that are available to alleviate the problems caused by breakdowns. Options include the use of:

- Observer Systems
  - Increased patrol frequency
  - Peak-period motorcycle patrol
  - CB radio monitoring
  - Aircraft
  - Transit ties
  - Volunteer observers
  - Professional observers

- Hardware Systems
  - Freeway telephone trouble number
  - Callboxes
  - Loop detectors
  - Closed-circuit television

- Service Patrols
  - Police service patrol
  - Dedicated freeeway patrol
  - Citizen service patrol
  - Highway agency service patrol

- Coordination Activities
  - Police/highway department coordination
  - Other public agency coordination
  - Citizens group liaison
  - Media ties to incident management agency
  - Private sector services coordination
  - Police or highway department/citizen group coordination
  - Media ties to citizens monitoring groups

- Incident Management — Field Procedures
  - Stationary response vehicle
  - Fast vehicle removal
  - Emergency lights policy
  - Wrecker contracts/agreements
  - Accident investigation sites
  - Response team

- Incident Management — Office Procedures
  - Dispatchers' manual
  - Hazardous materials manual
  - Alternate route planning
  - Communications training
  - Information digest

These options are all the potential ways that an engineer can reduce the impact of freeway incidents. Some of these options require cooperation with other agencies and if these options are chosen, agreements must be made between the agencies at an administrative level. Interagency cooperation is discussed in volume 1 and reviewed in chapter 16 of this volume.

The following sections expand on the options listed earlier.

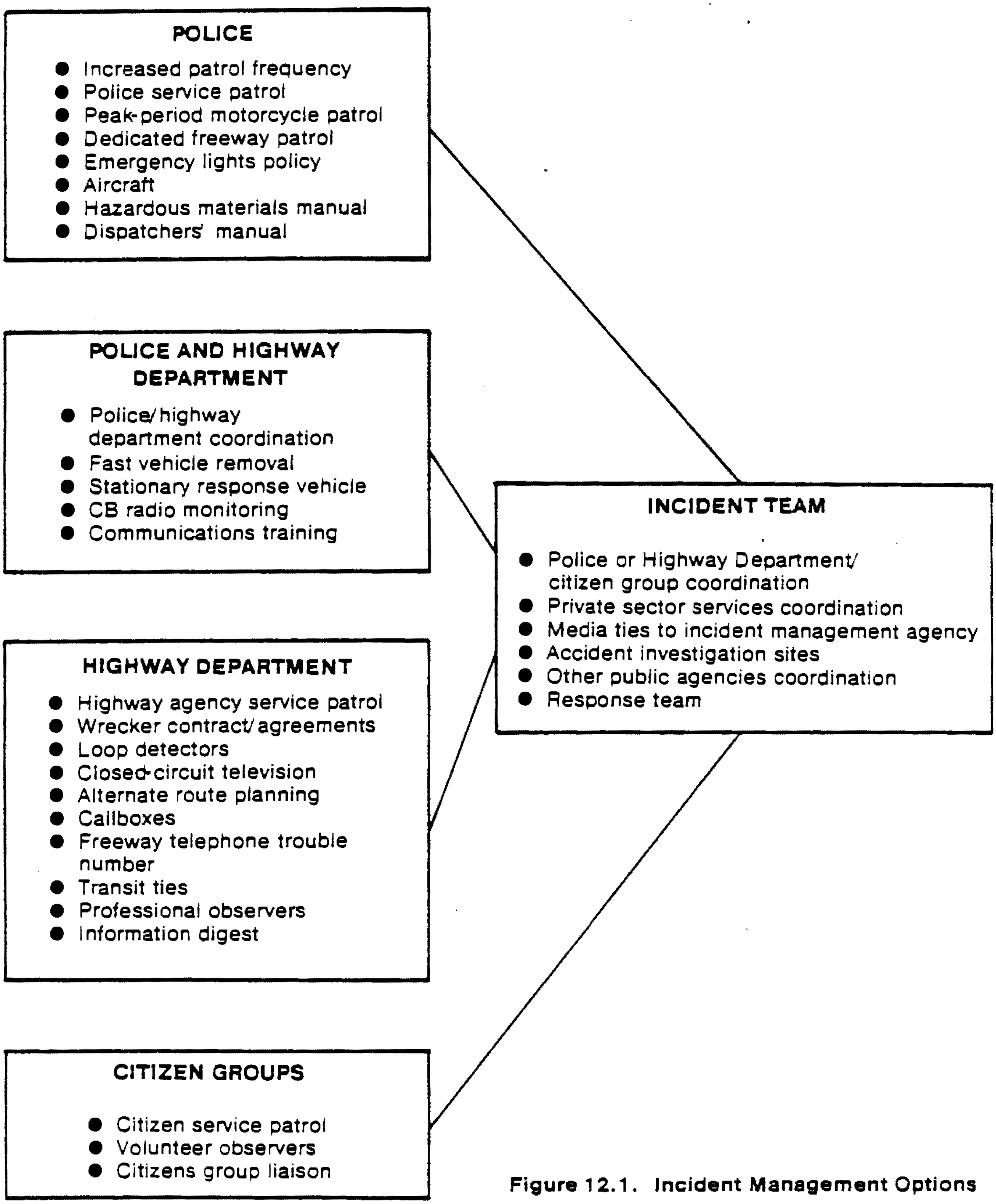


Figure 12.1. Incident Management Options

OBSERVER SYSTEMS

Increased Patrol Frequency

Implementation of this option involves increasing freeway patrol frequency by adding more patrol units or reducing the non-freeway duties of existing patrol units(2). More units can be added by acquiring new vehicles and men or by reassigning existing units to the freeway patrol duty. Reducing the non-freeway duties of patrol units requires a reordering of priorities or a new delineation of patrol beat boundaries.

Peak-Period Motorcycle Patrol

Implementation of this option is designed to take advantage of a motorcycle's maneuverability in congested traffic conditions. It involves procurement of motorcycles, training of drivers, and specification of operating procedures. The motorcycles must be able to operate at freeway speeds, a requirement which necessitates the procurement of motorcycles with an engine displacement of 900 cubic centimeters or greater.

CB Radio Monitoring

This option involves installing CB radios in police vehicles, service patrols, or a central control room. Implementation of this option requires a Federal Communications Commission (FCC) license, CB equipment, interagency operating procedures, and a publicity campaign to inform the public of the program.

An interesting and useful concept in CB monitoring is in use in Detroit where several data transmission outstations have CB antennas that transmit the signal to a central control room. Thus, the central operator can monitor CB broadcasts in a number of remote locations.

Any agency wishing to use CB radio as a means of incident detection must first apply to the FCC for an operating license. During 1976, many changes were made to the rules and regulations governing CB radio use. It is anticipated that these rules will continue to change. Since the current set of rules may soon be outdated, it is not explained here. The reader is advised to purchase the current version of Part 95: Citizen's Radio Service, Rules and Regulations from the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402.

It is desirable to inform the CB public of the fact that police vehicles are CB-equipped. It is particularly important for out-of-town motorists and in areas where officers are allowed only to monitor and not to transmit.

Aircraft

This option involves using aircraft primarily as incident detectors. Typical scenarios for implementing this option include police-owned aircraft, commercially-leased aircraft, and aircraft used cooperatively by the police department and a commercial radio station. In addition, guidelines for determining whether a police officer/observer will be used and other operating procedures must be developed.

If a helicopter is used, specifications similar to those provided in volume 4, appendix A, of reference 1 should be adopted. This specification is based on reference 3.

An operating procedure that is locale-specific must be developed. The procedure should include hours of operation on station, manning requirements, travel requirements, standby requirements, procedures for flying with ground patrols, and the administrative procedures typical of any operation.

If the aircraft is to be used for any purpose in addition to detection, one salient feature of the operating procedure that must be considered is response time. If the aircraft is to be used for managing onsite activities at major incidents, the standby location of the aircraft must be such that response time is reasonably short (5 to 10 minutes). Using the aircraft in this capacity requires the development of a prioritized procedures guide that specifies the data necessary for making onsite decisions regarding route diversion (see option, alternate route planning).

Another aspect that must be considered is non-patrol time necessitated by weather, repairs, and the performance of data collection activities for major incidents. The weather problem must be recognized as an operating handicap of this option, and operating procedures must be developed to increase the number of ground patrols during inclement weather. A similar redeployment procedure will have to be implemented when the aircraft is participating in major incident diversion activities.

<u>Transit Ties</u>

This option is applicable only when buses are radio-equipped and have significant route mileage on the freeway system. The ultimate aim of this option is to use these radio-equipped buses as additional detectors for the police agency. Implementation of this option requires that the bus drivers have training and/or operating procedures for reporting incidents to their dispatcher. In addition, a communications link between the transit agency communications center and the police dispatcher is necessary.

The transportation department's role in this option is highly dependent on which agency manages the transit operation. If the transit operation is privately run, the department's role is that of the intermediary between the transit agency and the police agency. If the transportation department provides transit service directly, it plays a support role, offering additional detection capabilities.

The technical expertise required of the engineer in a transportation department will also depend on the department's role. It may be necessary for him to become heavily involved in establishing the reporting points for each bus route, or he may be superficially involved and simply bring the two agencies together at a meeting.

Another aspect of the engineer's role may include convincing the transit operator that incident reporting is worth his commitment. The tactic taken to satisfy this concern will vary from locality to locality. However, the concept of an exchange of information, whereby the transit operator exchanges incident information for advanced warning of construction and other events on bus routes that will affect his schedule, is one important selling point. A second point can be made by using the planning procedures to estimate how the exchange of information will affect the transit operator in terms of reducing the delay on freeway routes. Data on lowered pollution rates and gasoline saved as a result of the information exchange can be developed and offered to the transit operator for use in publicity campaigns. As the bus drivers play a major role in this option, any trade union representatives should be involved in negotiations.

<u>Volunteer Observers</u>

This option involves using volunteer observers located in office and apartment buildings overlooking the freeway to detect incidents. The actions involved in implementing this option include recruiting observers, establishing an operating procedure, and providing an unlisted police telephone number.

Recruiting observers depends largely on the creativity and imagination of the citizen group or citizen organizing the observation effort. One attractive aspect of the qualifications of observers is that they only need to be able to see well with the aid of binoculars and to be able to use a telephone. These minimum qualifications suggest that almost anyone, including the physically handicapped, the elderly, and the retired, are potential candidates.

Any operating procedure that is developed should be simple and straightforward. An observer checklist similar to figure 12.2 would be adequate. To facilitate this information-gathering effort and to minimize relying upon the observer's memory, a question list should be developed in such a manner as to permit the information to be recorded directly beneath the respective question. Giving the observers multiple questionnaire forms or inserting a single form beneath a plastic overlay and using a grease pencil to record the information will achieve this objective.

Professional Observers

This option involves placing observers at strategic locations along a point facility to detect and aid with incident removal. Implementation of this option requires developing a position description and a set of operating procedures governing the observer's activities and prioritizing observer actions.

The location of an observer is dependent on the characteristics of the point facility. Obviously, a facility that houses an observer should be located so that sight distance is maximized. However, it is likely that such structures will be built for other purposes (e.g., toll collection, bridge tender) as well. In such instances, the planner should compare the benefits (as determined by the planning procedures) of locating the observation station to maximize sight distance for incident detection with the benefits of alternative locations for other uses.

HARDWARE SYSTEMS

Freeway Telephone Trouble Number

This option involves establishing a single, toll-free telephone number at existing telephones within the freeway corridor. This number is tied to the police dispatch center. Procuring this option is dependent upon the sophistication of local telephone company equipment. If the equipment is old, it may be technically impossible to obtain a fully toll-free number. Under such circumstances, the most desirable action is to obtain an easily remembered number, such as TRAFFIC (872-3342), and to ensure that sufficient lines are available so that the reporting motorist reaches the dispatcher easily. If the local telephone utility has relatively new equipment, it is likely that the area already has a 911 emergency number. If the 911 number qualifications are modified to include incident calls, sufficient line capacity must be added so that the emergency nature of the number is not degraded.

The optimal location for the terminus is the police communications center, preferably the dispatching area. This location is reasonable from a practical point of view, but to ensure that the full benefits of the system are realized, response to the reported incident must be prompt. Locating the freeway trouble number terminus in the dispatch center minimizes communication time and maximizes the potential for a prompt response.

Callboxes

Implementing the callbox detection option is a four-step process involving specifications, installation, monitoring, and maintenance of a callbox system. The first two steps require a level of specialized knowledge that is infrequently housed in highway departments or transportation

FIGURE 12.2. OBSERVER CHECKLIST

IN-LANE INCIDENTS

- What is the exact location (direction, route, cross streets)? ____
- Do there appear to be any injuries? ____
- How many vehicles involved? What types of vehicles? If a tanker was involved can you see a spill? Can you see smoke? What color is the smoke? ____
- What lanes are involved? (Which lane if only one lane?) Can you see the driver? ____
- Are the police on the scene? ____
- What is the traffic impact? How far is the traffic backed up? ____

SHOULDER INCIDENTS

- What is the exact location (direction, route, cross streets)? ____
- Do there appear to be any injuries? ____
- How many vehicles involved? What types of vehicles? If a tanker was involved can you see a spill? Can you see smoke? What color is the smoke? ____
- What lanes are involved? (Which lane if only one lane?) Can you see the driver? ____
- Are the police on the scene? ____
- What is the traffic impact? How far is the traffic backed up? ____

DEBRIS

- Is it in the roadway? Exact location? ____
- What lane(s) is the debris blocking? ____
- What kind of debris is it? ____
- Describe in as much detail as possible. ____
- Describe how it is affecting the traffic. ____

WIRES DOWN

- What is the exact location? Are the wires on the roadway? ____
- Is it causing a blackout? If so, what is the extent of the blackout? ____
- Can you tell if the wires are sparking? ____
- Are surface street traffic signals affected? ____
- What is this doing to traffic? ____

ANIMALS

In the event that a live or dead animal on the freeway right-of-way, determine the proper local authority to notify. (Enter that information here.) Specify the type of animal and the hazard (if any) it is causing.

PAVEMENT CONDITION

For snow or icy spots notify the police, DOT, or public works department of the condition. Determine this locally.

agencies. The third step generally involves an interagency agreement between the funding agency and the responding agency. The last step—maintenance—may be negotiated in an interagency agreement or in a contract with the system's manufacturer.

Specifying a callbox system is a detailed and lengthy process that is dependent on a variety of factors. Because these factors include the physical characteristics of the freeway system (e.g., interchange spacing, shoulder geometrics), traffic characteristics (e.g., volume, vehicle mix), and variations in callbox types (e.g., hardwire, radio, pushbutton, two-way) as well, no attempt will be made here to give sample specifications. Instead, it is recommended that the incident management system engineer refer to three FHWA documents that discuss issues associated with callboxes:

- Guide to Highway Communication Systems and Technology Design (L. Saxton et al., Report No. FHWA-RD-75-101, January 1975)

- Federal-Aid Highway Program Manual, volume 6, chapter 8, section 3, subsection 3

- Motorist-Aid Systems Study — State of the Art Report, available from the FHWA Office of Development Implementation Division (HDV-21), Washington, D.C. 20590

After consulting these references, if the engineer still desires assistance in wiring specifications, a qualified engineering or communications firm should be hired.

Loop Detectors

Loop detectors are not usually installed except as part of a complete surveillance and control system. In certain situations, however, their use may be a highly cost-effective alternative at high-accident locations. Detecting incidents with electronic loop detectors implies that some type of transducer buried in the roadway (usually a loop of wire 6 feet by 6 feet) transmits a signal to a roadside terminal which is connected by some means to a central computer (normally housed at the operations center of the point facility). An algorithm in the computer interprets the signal and outputs the result on some type of device (usually a printer and/or map board). Each of these components must be specified in such a manner as to be compatible with the others. Operating procedures for reacting to the output must also be developed. The use of detector-based surveillance for non-recurring congestion is discussed in chapter 10 of this volume.

Closed-Circuit Television

This option involves using closed-circuit television to detect incidents on a point facility. Although television is not normally considered a low-cost surveillance technique, under some circumstances it may be a viable alternative. Implementation requires a working knowledge of the state-of-the-art of closed-circuit television equipment. With such knowledge, detailed specifications to account for site-specific problems can then be developed. Volume 3 describes in detail the use of closed-circuit television systems on freeways.

SERVICE PATROLS

Police Service Patrol

This option consists of supplementing the existing police patrol with police-operated tow trucks, pickup trucks, or station wagons. Implementation of a tow truck police service patrol requires specification of a medium-duty wrecker, operator's position description/qualifications, guidelines for operating the wrecker patrol, and a means of keeping the operating procedures current. Use of a police-operated pickup truck or station wagon would be implemented in a manner analogous to that outlined in the description of the highway agency service patrol option.

The specifications for a wrecker vehicle that will enable the full impact of delay reduction to be achieved are given in volume 4, appendix A, of reference 1. This specification is derived from material in reference 4 and discussed in volume 3.

Dedicated Freeway Patrol

The concept of a dedicated freeway patrol consists of equipping the ordinary patrol vehicle in such a manner so as to be more useful for incident management procedures. Major operating equipment consists of a gasoline transfer device, a push bumper, a rollaway dolly, and a multipurpose jack. Other useful equipment includes water, oil, jumper cables, and certain smaller tools.

The operation of a gas transfer device involves turning a petcock while the engine of the police vehicle is running. One gallon of gas is usually sufficient to start most vehicles. It may be necessary to prime the engine if the vehicle does not start after several attempts. An alternative to priming the engine is to carry pressurized cans of commercially available "quick-start," which is commonly used to aid starting in cold weather. The advantage of using this spray is that it is much less time-consuming than priming the engine. Disadvantages are that it is considerably more expensive than priming with gasoline and storing this highly flammable substance in a trunk compartment with flares represents a substantial fire hazard.

Citizen Service Patrol

This option involves patrolling the freeway in a service vehicle owned, operated, and equipped by a group of citizens. Such a patrol may be sponsored in a variety of ways; for example, by citizen groups, auto dealerships (as a civic action), or other public-service groups. The equipment necessary to implement this option is a vehicle equipped to service several types of incidents. Development of both an agreement with the police and an operations procedure or guideline is necessary, as is a general understanding of the local laws governing citizen-given aid on the freeway.

Either a formal or an informal agreement with the local police should be established. The purpose of such an agreement is twofold. First, a determination of the legal requirements of the patrol will be made. For example, it may be necessary to equip the vehicle with special warning lights. Second, the police must be informed of this activity because the patrol will be operating on a facility on which the police are legally responsible for enforcing laws and regulations. The agreement need not be formal but should touch upon every aspect of the freeway environment that will be affected by the citizen patrol.

Highway Agency Service Patrol

Implementation of an agency service patrol should begin with police department review and approval of an operations plan. Police department concurrence in an agency plan to operate a service patrol is often required because of the police department's traditional jurisdiction over traffic in the freeway environment. This responsibility has led to the development of a working relationship with private service station and wrecker operators who would be directly affected by a public agency service patrol. In the interest of maintaining their relationship with that group, the police may request that agency assistance be limited in some manner so as not to infringe on private business.

Conversely, the police may be dissatisfied with private operator performance and welcome an agency proposal to assume both service and towing responsibilities. In either case, an interagency agreement should be formulated clarifying each agency's role and their mutual relationship with regard to vehicles in need of service.

The service patrol's communications capability is another factor that may be affected by an interagency agreement between the highway agency and the police agency. Established highway department radio frequencies can adequately handle the transmissions of a new unit. To improve service, however, the patrol can be equipped with a police radio. Vehicles with this dual communications capability can reduce response time by receiving direct messages from police units rather than those passed from dispatcher to dispatcher. If their use of a police radio is approved, they can also report detected accidents directly to the police. Equipping patrol vehicles with CB radios might be considered, but it should be recognized that it is difficult to listen to a CB radio and be responsible for other agency transmissions at the same time.

COORDINATION ACTIVITIES

Police/Highway Department Coordination

Because this option involves all levels of each agency's structure, a commitment to interagency cooperation must originate at the top with the chief administrator or executive. The police or highway department engineer should start by gathering evidence of latent problems, formulating suggestions for improvement, and presenting the findings to the chief administrator. The planner should not wait until the occurrence of a serious incident exposes the existing system's shortcomings and forces recognition of the need for closer interdepartmental ties. Reminders of poorly managed incidents in the past, and of the potential for improved publicity in the future, will reinforce the engineer's advocacy position.

The purpose of this option is to develop a sound working relationship between two public agencies at all levels. This effort is initiated by securing the chief administrators' endorsements of a cooperative, interagency approach to freeway operations and by arranging meetings of administrative and supervisory personnel from each agency. At these meetings, police and highway department responsibilities for incident management should be identified and discussed. Then, a joint operational policy statement should be drawn up and signed by each agency's chief administrator. Distribution of this agreement throughout all organizational levels is the next step in the implementation process. Finally, a monitoring procedure should be established to ensure that the continuing benefit of the cooperative relationship is realized.

Although police and highway department responsibilities for freeway operations are generally well established and mutually exclusive, there is often some overlap regarding freeway incident management. Notification procedures, vehicle removal, and communication with the media are examples of the types of issues that require clarification. In particular, onsite authority deserves special attention. Consideration should be given to resolving this question by establishing a command post at major incidents. A command post consists of a specially marked police patrol vehicle from which all onsite incident management activities are coordinated. All public agency personnel, including highway department employees, are required to report to the command post upon arrival at the incident scene. At the command post, cleanup activities are planned jointly by the ranking representatives of each agency. In addition, all communications with the central dispatching facility originate from the command post.

Other Public Agency Coordination

The purpose of this option is to develop working relationships between the agency having primary incident management responsibility and other public agencies or public-spirited organizations that perform specialized or back-up incident management functions. Implementation of this option requires developing: first, realistic goals and procedures for response and onsite activities with those agencies that are commonly requested to respond to incidents; and second, notification and response procedures with those agencies that are only required for major or rare incidents.

Agencies that are commonly needed at the incident scene include fire/response services, ambulance services, and certain utilities. It is often necessary to make these agencies aware of the impact that their actions may have upon traffic at an incident site. Typical problems that can occur and must be resolved include:

- Parking emergency vehicles in lanes that need not be closed to traffic
- Inefficient distribution of emergency equipment at the incident site, which hampers emergency personnel from effectively performing functions
- Responding with more vehicles than are necessary and failing to remove unnecessary vehicles from the site immediately
- Failing to report to the command post, as discussed in the police/highway department coordination option

Agencies that are infrequently called to incidents include the state environmental protection agency, the agency responsible for hazardous materials spills, Federal agencies, and others. The primary need is to develop a means of assuring prompt notification of the proper agency and rapid response by its personnel.

Citizens Group Liaison

Implementation of this option involves the formation of a relationship between the police or highway department and citizen volunteers who detect and report incidents. These volunteers are generally mobile CB radio operators who belong to existing monitoring clubs or who join newly formed clubs. Thus, the first step is to contact local CB clubs, or, in areas with no existing clubs, to encourage the organization of a club. A system for the direct reporting to the public agency of citizen-observed crime, incidents, and road conditions is then established. Such a system can consist of a direct CB radio communication or the transfer of information via landline. Next, a police or highway department employee assigned to represent the department within the CB club should train members in proper reporting and communication techniques. Public agency influence over the reporting system should be further maintained by distributing club decals and membership cards selectively or in the case of a newly organized and sponsored club, by controlling membership.

Media Ties to Incident Management Agency

The intent of this option is to establish a formal relationship with commercial radio stations to ensure that the incident information transmitted to the public is timely and correct. Such a relationship can be formalized with a contract agreement, or memorandum of understanding.

The essence of this option is that the information supplied by the freeway incident management agency be broadcast within a specified time period (e.g., 5 minutes) after receipt. Of utmost importance are the accuracy and clarity of the information reported to the radio station. If the information is not broadcast directly as received from the freeway incident management agency, a determination must be made regarding the correctness of what was actually broadcast.

Private Sector Services Coordination

Implementation of this option involves developing working relationships with private companies to supply specialized equipment needed in extraordinary incident situations. First, an inventory of existing public agency resources must be conducted. This process leads directly to the next step, assessing what additional equipment and services are needed to respond effectively to unusual incidents. Next, private suppliers of these specialized services must be located, and their commitments to provide equipment and/or service obtained. These commitments are then made definite by negotiating agreements or contracts concerning the conditions under which private sector resources will operate, their costs, and the method of compensation. Finally, a notification procedure for dispatching private sector services should be formulated jointly.

The first step in determining what specialized equipment and vehicles are to be furnished by private suppliers is to inventory the resources that are currently available to the engineer's agency either from its own vehicle/equipment pool or from those of other public agencies. A list of these resources, which are commonly required to clean-up incidents involving overturned trucks and spilled cargoes, is presented in volume 3. The public agencies that should be contacted regarding the availability of these resources are also listed in volume 3.

Police or Highway Department/ Citizen Group Coordination

This option involves the development of relationships between citizen groups concerned with incident management and the various police and highway agencies responsible for actual incident management in the area. The purpose is to adopt procedures for making maximum use of citizen efforts in incident management activities. To implement this option, the citizen group must first determine the needs of the public agencies and then convince official representatives that the citizen group can make a significant contribution. Once these relationships are established, detailed internal operating practices must be developed.

Media Ties to Citizens Monitoring Groups

Implementing this option involves establishing a relationship between the citizen group and the local media, probably a radio station. The activities necessary to implement this option are an organized citizen effort to collect incident information, a means of interpreting the information and preparing it for broadcast, and an agreement with the radio station.

Developing an organized citizen effort implies that the citizen group has successfully implemented CB radio monitoring, volunteer observers, or a citizen service patrol. In effect, this means that the group has information that may be of use to the radio station.

INCIDENT MANAGEMENT — FIELD PROCEDURES

Stationary Response Vehicle

Implementation of this option requires analysis of historical incident data, specification of the response vehicles' detection responsibilities and communciation capabilities, preplanning of access routes and procedures, and documentation of standard operating procedures.

Analyzing historical incident data for a point facility or a freeway system enables the planner to determine the optimal type of response vehicle and its deployment. Generally, the vehicle will be a wrecker, since it can service and/or remove the majority of lane-blocking incidents. Point facilities are likely to require permanent, full-time deployment of an emergency vehicle. An entire freeway system, however, may only require peak-period deployment at high-accident locations.

Locating response vehicles near high-incident locations or point facilities is of little benefit unless incident detection is achieved quickly and the response vehicle operators are notified. Response time is minimized if the operators themselves have detection responsibility. Depending on local topography, this can be accomplished by using binoculars or by monitoring a CB radio. Alternatively, the response vehicles can be equipped with radios operating on the highway department, police, or point facility frequency. To ensure that the response vehicles can reach any incident within their defined area of operations, access routes and procedures must be preplanned.

Fast Vehicle Removal

The fast vehicle removal option involves methods to ensure that disabled, abandoned, or damaged vehicles do not constitute hazards to other motorists. There are basically two variations of a fast vehicle removal policy, each with its own method of implementation. First, there is fast vehicle removal viewed as the responsibility of the motorist involved. Implementation of this variation may require a lobbying effort to create new legislation, a publicity campaign to announce the new policy, translation of the policy into operating procedures, and an enforcement effort to cite violators and effect removal if the motorist fails to take the initiative.

Depending on existing municipal ordinances and state laws, new legislation may or may not be necessary to force public initiative in vehicle removal. Laws may already exist that can be interpreted as requiring motorists to arrange for vehicle removal, although such laws are likely to be somewhat vague about what constitutes a "disabled" vehicle or a "reasonable" amount of time for effecting removal. If more specific regulations are sought (for example, the declaration of vehicles with flat tires to be drivable and not disabled, or the establishment of a maximum time limit for leaving vehicles in non-hazardous locations), a lobbying effort for the enactment of new legislation is required. In either case, the engineer should consult the agency's general counsel for information on the current legal situation and suggestions on proposing new legislation, if necessary.

The other fast vehicle removal option consists of public agency initiatives (usually taken by the police or highway department) in removing disabled, abandoned, or damaged vehicles. This variation involves the use of specialized removal equipment and possibly, the enactment of legislation protecting the public employees who use it.

Legislation absolving public employees of liability for damages in removing disabled vehicles should exist if they are to assume an active role. Some agencies may be able to justify fast removal activities under their current mandate, but the general counsel must make that determination. Adequate control of vehicle removal by public employees can be maintained by vesting authority only in law enforcement officers or others acting with their permission.

Emergency Lights Policy

The purpose of this option is to develop and adopt a standard policy regarding the use of emergency lights (generally, four-way flashers, rear deck, and rotating rooftop lights) on public agency response vehicles. The development phase involves a review of existing policies, formulation of recommended changes, and approval of the new guidelines by agency administrators. The adoption phase requires a revision of operating procedures and retraining of field personnel. In some areas, these steps may be followed by removal and replacement of certain flashing lights currently mounted on response vehicles.

The engineer should first review the department's formal or informal emergency lights policy to see if it considers the trade-off between warning oncoming motorists and attracting undue attention to an incident. Awareness of this issue should be reflected in a procedure that is based on the location of the incident (in lane or on the shoulder). When an incident obstructs a traffic lane, emergency lights should be used as a warning signal to oncoming motorists. When an incident is located on the shoulder, however, no lights should be used, since they attract attention and lead to rubbernecking delays.

Wrecker Contracts/Agreements

It is presumed that existing wrecker service is provided on an ad hoc basis and/or that the existing method of obtaining wrecker service is unsatisfactory. Implementation of this option requires significant legal activity, which may receive opposition from existing wrecker operators. These legal requirements must be formulated properly so that existing services are not put out of business but yet are regulated to the extent required to obtain the minimum response time and the delay savings indicated in the planning estimates. Accomplishment of this goal may require not only the elements discussed in the following sections, but also a public relations/information effort aimed at all of the wrecker operators and at the public at large.

To establish this option, it may be necessary to enact an ordinance that gives the police authority powers to establish wrecker contracts, lay the ground rules for such contracts, and determine the penalities for the non-compliance of wrecking companies.

Three types of contractual agreements can be considered:

1. Rotation list

2. Bid contract

3. Consortium bid contract

With the rotation list procedure, the dispatcher maintains an indexed card file of eligible operators for each sector. As each tow operator is called, his card is placed at the rear of the file. This procedure is formalized via a standard operating procedure or memorandum.

With the bid contract procedure, the incident management area must be divided into sections that can be served logically and profitably by one operator meeting the response time requirements. From the standpoint of the public agency, it is desirable to have only one contractor. However, it may be impossible for one contractor to service the agency's entire jurisdiction adequately. Under such circumstances, several operators may choose to form a consortium. This allows a single operator to have one or more subcontractors. Thus, for any incident within the particular area, the police still have only one number to call. The only police recordkeeping that is done is the validation of whether the predetermined response time is being met.

Accident Investigation Sites

Implementing the accident investigation site option(5) involves imagination and engineering judgment in locating accident investigation sites. If possible, these sites should be located out-of-sight of freeway drivers so that rubbernecking is minimized. A site could be nothing more than a sign indicating that a portion of the shoulder or an off-ramp is the accident investigation site. On the other hand, the site could be completely paved and located beneath an overpass. Whatever type of site is used, provisions should be made for parking at least five vehicles: a police car, two disabled vehicles, and two wreckers. Ideally, to reduce the number of sites required, sites should be located at each interchange in such a manner as to allow both directions of traffic to use each site. However, such a location is dependent upon the type and spacing of the interchanges.

To ensure that the sites are used, a commitment from the police agency is required. In addition, the location of each accident investigation site must be transmitted to each trooper and, if feasible, to each tow operator.

Response Team

The response team option consists of creating a team of traffic engineers capable of responding to major traffic incidents for the purpose of implementing the alternate route planning option, major traffic diversions, or other onsite management and traffic control activities. Implementation of this option requires selecting team members, developing procedures to activate the response team, and equipping the response team.

Team members should be experienced traffic engineers, should have a working knowledge of the freeway and all secondary routes, should have played a major role in developing the alternate routes, must be familiar with all of the diversion techniques, and be at least casually familiar with hazardous materials handling as indicated earlier.

An operating procedure is required to enable the dispatcher, in consultation with the beat or watch commander, to determine when to call out the response team. The essential feature of the operating procedure includes estimating the duration of the incident and the maximum queue length likely to occur. Methods for estimating delay are discussed in detail in chapter 3.

INCIDENT MANAGEMENT – OFFICE PROCEDURES

Dispatchers' Manual

Implementation of this option involves collecting information specific to incident management and organizing it for use by the dispatcher. This is an important element of all options since the dispatcher is a common central contact point. Detailed guidelines are discussed in volume 3. In addition, sample formats for organizing incident management information are presented.

Hazardous Materials Manual

Implementation of this option involves the preparation of three documents concerning handling hazardous materials at the incident site. The first document is a glove-compartment-sized guide to hazardous cargo placarding, labeling, and site stabilization. This guide is to be used by police or highway personnel who are generally the first emergency workers to reach an incident. The second document is a more comprehensive manual for use by the central dispatcher. It contains both information about the toxic characteristics and flammability of an identified hazardous material, and precautionary, first aid, and firefighting measures. The third document, designed for use by firefighting units, cleanup crews, and special hazardous materials-handling teams, is a comprehensive document listing the chemical properties, recommended cleanup procedures, and appropriate first aid measures to be taken in case of the spillage of a wide variety of chemical compounds. Volume 3 contains a detailed discussion of these three types of manuals and presents illustrative examples of each.

Alternate Route Planning

Implementation of this option involves developing pre-incident, detailed alternate route contingency plans for different levels of freeway incidents any place on the freeway system. The first step is to develop a priority list of the order in which the alternate routes should be developed. The second step is to inventory all existing surface streets that might serve as an alternate route for every section of freeway. This effort leads into the third step, expressing the alternate route(s) capacity in terms of the amount of diverted traffic that the secondary route is able to serve. Next, every local agency that has jurisdiction over the surface streets of the proposed alternate routes must be contacted, and its coordination and/or concurrence obtained. Finally, maps indicating the alternate route(s) for each section of the freeway should be developed.

It is necessary to assign priorities to the alternate routes under consideration because comprehensive planning is a large and time-consuming task. It is recommended that major routes have alternate routes defined as a first priority. Identification and selection of alternate routes should consider the presence of:

- High-accident locations
- Sensitive locations (e.g., hospitals, schools)
- Route paths that would be unclear, confusing, and/or difficult for the average motorist to follow
- The easiest way by which the motorist can return to the major route

The freeway(6) must be divided into sections (e.g., between successive exits), and all surface streets that could serve as alternate routes must be assessed in terms of capacity (including grades, width, curvature, surface condition, turning radii, number of intersections, and side friction). From this analysis, potential primary and secondary alternate routes can be developed.

The capacities of the primary and secondary alternate routes are then converted into a percentage of the freeway capacity; that is, an alternate route may have a capacity equivalent to a freeway with two lanes blocked during the rush hour, yet able to meet the demand on the freeway at other times. Thus, for a full freeway closure during the rush period, both primary and secondary routes may have to be used. This information will then be arrayed in matrix format and organized by time of day versus severity of the incident (in terms of estimated duration and number of freeway lanes blocked), which indicates various stages of the alternate route implementation process.

After the initial potential alternates have been developed, each local agency having jurisdiction over the surface streets to be used must be contacted and brought into the planning process. Their working knowledge of the facilities must be incorporated into the development of the final plans for the alternate routes. Finally, each agency must agree to the role it will take in the event that a diversion is necessary.

All of the above planning must be condensed into a usable operational procedure guide. The most useful means of conveying this information is through a series of maps. Another item that might be shown on the maps is the number of policemen, maintenance men, engineers, others, and equipment that would be required to implement a full-scale diversion. Consideration should also be given to preparing a schematic map that is easier to read than a conventional road map.

Communications Training

This option involves the instruction of telephone and radio operators in proper voice communication techniques for reporting and receiving information concerning freeway incidents. This training is meant to supplement existing training programs. The techniques include following the FCC regulations that govern radio transmissions; using standard reporting formats to transmit accurate, complete, and succinct messages; and practicing good voice quality and enunciation so that messages are intelligible. These techniques should be taught to base station dispatchers in particular, and to all people who engage in incident-related communications, from civilian CB radio operators to police and highway department personnel. The purpose of these techniques is to improve the quality of detection reports and dispatching messages and to minimize communication time.

For public agency radio operators, this option is implemented by supplementing standard communications training (e.g., the use of radio equipment, logkeeping, and recordkeeping) with instruction in the use of voice communication techniques. The proper use of mnemonic devices, the Ten Signal Aural Brevity Code, the international phonetic alphabet, and military time reporting can be taught with written materials from an operating procedures manual and with oral lessons using tape-recording equipment to allow radio operators to hear their own transmitted words. An excellent source of material for an operating manual is The Public Safety Communications Standard Operating Procedure Manual, published by Associated Public-Safety Communications Officers, Inc. In addition, public agency radio operators should be familiarized with the FCC regulations governing their radio system. These can be found in title 47 of the Code of Federal Regulations, chapter 1, subchapter D, part 89, "Public Safety Radio Services."

Civilian CB radio operators can also be instructed in the use of these communication techniques. This is best accomplished by contacting local CB radio clubs and monitoring organizations such as REACT and ALERT. Training sessions led by a police or highway department representative should be offered in conjunction with membership in public-agency-sponsored citizen participation projects.

Information Digest

The information digest is a resource document compiled by a CB-monitoring club for primarily assisting out-of-town mobile CB operators. Its incident management function is to enable the CB monitor to assist the CB user who has detected or been involved in an incident. In addition, it serves as a motorist-aid system. It consists of information in the form of local maps, telephone and address lists of institutions, agencies, tourist attractions, and businesses whose services the CB motorist may require during his trip. Special notation is given to 24-hour establishments, since they are often difficult for a stranger to find during the night or early morning.

REFERENCES

1. Peat, Marwick, Mitchell & Co. "Alternative Surveillance Concepts and Methods for Freeway Incident Management." Prepared for the FHWA RD-77-58 through 63, inclusive.

2. Lu, C. K. and May, A.D. Analysis and Design of Freeway Incident Response Systems, DOT-OS-40014, Interim Report, University of California at Berkeley, Institute of Transportation and Traffic Engineering. November 1974.

3. Juge, J. D., et al. Early Detection and Rapid Removal of Disabled Vehicles and Other Hazards from the Freeway, vol. III: "Airborne Television System; Incident Management Team." NHTSA Project DC 69-001, Final Report, Washington, D.C., June 1974.

4. Juge, J. D., Kennedy, K. R., and Wang, T.C. Early Detection and Rapid Removal of Disabled Vehicles and Other Hazards from the Freeway, vol. II: "Rapid Removal; Emergency Tow Service." NHTSA Project DC 69-001, Final Report, Washington, D.C., June 1974.

5. Texas Transportation Institute. "A Study of Accident Investigation Sites on the Gulf Freeway." NTIS: PB223583/AS. August 1972.

6. Adapted from a presentation by D. H. Roper to the meeting of the Highway Research Board Freeway Operations Committee, A3EOE, in Columbus, Ohio, June 27 and 28, 1973.

volving charge account at the department store. But when he is forced to maintain these debts with high interest rates in order to pay his taxes it is logical for him to want his tax money invested in public works to earn returns on his investment more nearly equivalent to his cost of money."

Another argument is that agencies are limited to expenditure in the current year and that if the proposed projects are to be financed from taxes raised in the current year, then a zero interest rate should be used. That is to say, expenditure n years from now will be paid from taxes collected during that year and, therefore, the opportunity to invest that money today does not exist. This argument seems particularly important in the case of a long-term staged construction or maintenance costs. A survey of states(3) indicates that 45 percent use a zero percent interest rate for all project types with the 89 percent using zero percent for maintenance costs.

However, if money has to be borrowed to finance a project, it would seem reasonable that the interest rate on private investments and public investments should be the same and current interest rates should be used. We do not recommend a specific interest rate for use, but suggest that the engineer investigate current rates used in other public projects, as well as current marketplace rates. A great deal of caution should be used when applying interest rates. The 1981 discount rate recommended by the Federal government was 10 percent.

Effect of Assumed Useful Life

In an economic study, capital investments are spread uniformly over the assumed useful life of each element in the highway improvement. Where salvage value is not considered, this uniform annual charge for principal and interest is found by multiplying the capital investment of the element by the capital recovery factor (CRF).

As the assumed useful life increases, the CRF approaches the interest rate. At high interest rates this convergence occurs rapidly; at low interest rates, this convergance occurs slowly. Therefore, economic studies involving higher interest rates are relatively less sensitive to changes in assumed life, and at low or zero interest rates the sensitivity is correspondingly higher. For example, at 7 percent the increase in the annual cost of capital recovery, when the assumed life is decreased from 30 to 20 years, is 17 percent. At zero percent the increase is 50 percent.

Effect of Assumed Salvage Value

The salvage value of a component part of the highway improvement is its dollar worth at the end of its assumed useful life. The salvage value assumed is dependent upon the assumed useful life. With realistic interest rates and relatively long component lives, the effect of salvage value is very small. Under these conditions the effect of including and excluding salvage value would result in an error of a magnitude that is considerably less than the expected error in other estimates. Therefore, for most highway facilities, or components of highway improvements, zero salvage value is appropriate in economic studies.

Effect of Assumed Average Daily Traffic (ADT)

Reductions in road-user costs provide one of the major justifications for highway improvements; therefore, they have a significant influence on an economic study. The annual savings in road-user costs is normally directly proportional to the estimated ADT for the analysis period. Projections of traffic on a new or

Identifying and Analyzing Problems	2. Data Collection 3. Delay Computation 4. Environmental Impacts 5. User Simulation Models
Procedures for Designing and Evaluating Solutions	6. Detection of Recurring Congestion 7. Ramp Control 8. Mainline and Corridor Control 9. Bus/Carpool Priority Control 10. Detection of Non-Recurring Congestion (Incidents) 11. Incident Detection Algorithms 12. Planning for Incident Response 13. Cost-Effectiveness Evaluation Techniques 14. Simulation Models in Design and Evaluation
Implementing Solutions	15. Financial Planning and Staged Construction 16. Interagency Cooperation

This chapter outlines the general procedure for evaluating freeway management options and describes the accounting methods used by engineers to compare the cost-effectiveness of these options. The effects of interest rates, useful life of the option, and other factors are also discussed.

Figure 13.1 outlines the procedure used for evaluating a series of options that may provide solutions to specific freeway management problems.

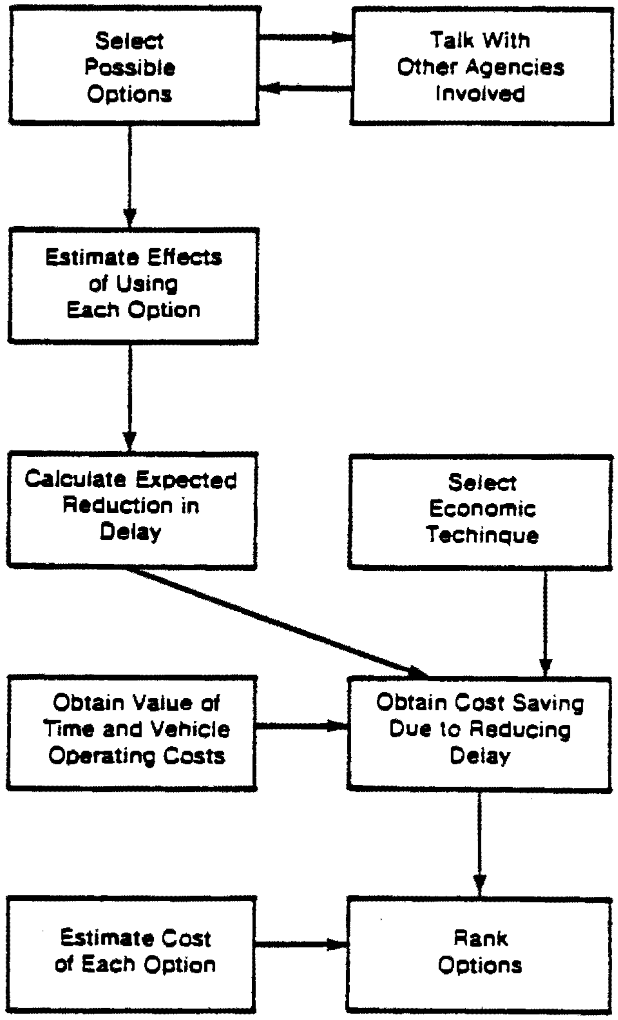


Figure 13.1. Evaluation Process

OVERVIEW OF EVALUATION PROCESS

Select Possible Options

A selection of options to be evaluated must be identified. Many options may not be suitable for particular freeway systems. It might be necessary to evaluate some options for particular parts of a freeway system and other options for the remainder of the system. For example, closed-circuit television might be a suitable option for that part of a freeway system near the central business district, while callboxes would be a more viable option on suburban and rural sections of the freeway.

During planning of a new freeway system or modification of an existing freeway, engineers and planners should select a series of options that would:

- Be locally acceptable
- Suit the existing or proposed system
- Provide a solution to an existing or potential problem

Talk With Other Agencies

Knowledge of local conditions provided by other agencies is an important part of this selection process. Many of the potential solutions to freeway problems involve the cooperation of other agencies, and it is advisable to discuss the types of solutions being considered with these agencies to obtain their opinions and, if possible, an esitmate of the extent and costs of their involvement. An example of this would be an option that increases the police patrol frequency on the sections of the freeway under management. It is likely that this option would require additional vehicles and radio equipment as well as a commitment from the police.

Estimate Effects of Using Each Option

To estimate the effects of each option being evaluated, specific information must be known and analyzed. For example, if a dedicated freeway patrol is the option under consideration, two factors must be known: the headway between patrols and the ability of patrols to detect incidents that have occurred in the other travel direction. Once identified, the time necessary for a dedicated patrol to detect an incident can be estimated.

Using an incident tree (see figure 2.5 in chapter 2) and the number of vehicle-miles per mile on the freeway, the number of incidents can be estimated. This is the value required for delay computation. However, the engineer must also estimate the time duration of the incident for the existing conditions in order to fully evaluate the effects of the option. On existing freeway systems this is not difficult as incidents will be occurring and surveying procedures will provide the relevant information. On a proposed facility, the detection response and clearance times of existing freeway systems should be examined in order to make estimates.

Calculate Expected Reduction in Delay

Delay is one of the best and most common measures of congestion, because it is cumulative. Delays at differing points on the freeway can simply be added to provide a total number of vehicle-hours of delay for each option.

The calculation of the expected reduction in delay is determined using the basic principal of evaluating the results of excessive demand. These calculations are straightforward for recurring congestion because the capacity of the location being analyzed remains constant. The calculation of delay for non-recurring congestion is more complex because the demand and capacity may both change, and the degree and duration of these periods of reduced capacity are dependent upon the type of incident, and the detection, response, and incident removal time of the freeway management system. Four generic types of equations have been developed to analyze various situations which cover: a simple reduction in capacity; a reduction in capacity and a period where all traffic is stopped to clear the incident; a reduction in capacity and a period when the capacity is between this reduced value and its normal value; and an incident in which the demand is decreased because of the diversion of traffic from the facility or the end of the peak period.

Obtain Value of Time and Vehicle Operating Costs

If the evaluation is to be performed in terms of vehicle-hours of delay alone, or it is an evaluation intended to reduce a specific problem (for example, accidents), then this step may be skipped. However, if the evaluation is to include a dollar value, the engineer must use a value of time and a vehicle operating cost. Vehicle operating costs can be obtained from the American Automobile Association and the American Trucking Association. Operating costs for 1981 were:

- Medium-sized autos: $0.25 per mile
- Trucks: $1.25 per mile

The values are subject to considerable fluctuation resulting from oil price changes and inflation. Thus, any values used should be current. The value of time is a contentious issue and there are many different techniques and methods used in its calculation. If any specific value has been used by the local agency, it would be reasonable to use this value. The 1980 value of time quoted by AASHTO is $3.00 per hour. This value encompasses work and pleasure trips as well as allowing for the non-working sections of the population.

Select Economic Technique

Engineering economic textbooks use four differing procedures for comparing alternatives in monetary terms:

- Annual cost method
- Present worth method, in which comparison is made on a present worth basis of all present and future expenditures
- Rate of return method, which involves the determination of the interest rate at which the alternatives are equally attractive
- Benefit-cost ratio method, developed by civil engineers involved in public works expenditure

These techniques are explained in more detail later in this chapter.

Obtain Cost Saving Due to Reducing Delay

This function involves taking the expected reduction in delay, separating the vehicle delay into truck and auto components, and then using vehicle operating costs and the value of time to put a dollar value on the savings due to reducing delay.

Estimate Cost of Each Option

The engineer and planner have to obtain the best estimate of the costs associated with the option under consideration. These, of course, depend upon the type of option and its degree of sophistication. Some of the options are very inexpensive as they merely involve initiating an agreement between two agencies. Others require administrative effort but no capital expenditure on hardware; an example of this would be the production of a hazardous materials manual.

Rank Options

In presenting a summary of freeway management operational options, costs and some traffic flow parameters should be included. Indeed, as mentioned earlier, if a specific problem is to be addressed, the evaluation can be limited to predicted before-and-after values and separate associated costs. The format of this presentation will, to a large extent, depend upon the cost-effectiveness evaluation technique involved.

ACCOUNTING METHODS(1)

Annual Cost Method

The annual cost of a capital investment to be recovered in n years with interest is found by multiplying its first cost by the appropriate capital recovery factor. With the uniform amount so determined, if charged at the end of each year for the assumed useful life, the initial investment will be repaid exactly, with interest.

The total annual cost of a highway improvement is the sum of all annual costs of capital recovery, plus the annual costs of maintenance and operation, plus the annual costs of the road users, plus the annual costs of traffic accidents. Annual costs are computed for the existing facility and for each of the proposals for improvement. The alternative resulting in the lowest annual cost represents the best solution from an economic standpoint.

This method has a serious drawback, in that widely differing results are obtained for various assumed interest rates. Generally in the annual cost method, low interest rates favor those alternatives that combine large capital investments with low maintenance and operation or user costs, whereas high interest rates favor reverse combinations.

Present Worth Method

In making comparisons by the present worth method, all costs and benefits are reduced to present-day values. One difficulty encountered in the application of present worth comparisons is that equal periods of service must be analyzed. The following problem illustrates this point.

Two proposed options are to be compared, using an interest rate of 6 percent. Option A has a first cost of $500 and an estimated life of 8 years. Option B has a first cost of $1,000 and an estimated life of 20 years. Annual operating expenditures for both options will be the same.

If a present worth comparison is to be made, the period of service under consideration must be the same for the two options. Therefore, it is necessary to consider enough renewals of each option to serve for the number of years which is the least common multiple of the estimated lives of the two structures, in this case it is 40 years. It is necessary to assume that renewal costs will be the same as the first cost (see table 13.1). Therefore, according to table 13.1, option A would be selected as the best alternative.

TABLE 13.1. ILLUSTRATION OF PRESENT WORTH METHOD

OPTION A

First cost	$ 500
Present worth of first renewal (500 x 0.6274)	314
Present worth of second renewal (500 x 0.3936)	197
Present worth of third renewal (500 x 0.2470)	124
Present worth of fourth renewal (500 x 0.1550)	78
Total present worth of cost of 40 years of service	$1,213

OPTION B

First cost	$1,000
Present worth of first renewal (1,000 x 0.3118)	312
Total present worth of cost of 40 years service	$1,312

In applying present worth to highway economy studies, comparison is made between the sum of the present worths of the benefits and costs for maintenance and operation during each year of the analysis period, and the total capital investment required to produce these benefits. This method tends to minimize the effects of large future benefits arising from projected increases in traffic volumes, and tends to place greater value upon benefits accruing in the near future.

Rate of Return Method

As explained above, the annual cost method of comparing alternatives involves converting each alternative into an equivalent uniform annual series of payments. This method is preceded by a decision on the appropriate interest rate of minimum attractive return to be used in the conversion.

Another method of comparing expenditure time alternatives is to find the interest rate that makes them equivalent to one another. In this method the interest rate is the unknown in the problem. In the common situations of comparing two alternatives, one with a lower capital investment and higher maintenance, operation, and road-user costs, this unknown interest rate may be thought of as the prospective rate of return on the excess of investment in the alternative with the higher capital investment.

Therefore, the rate of return method involves determining the rate of interest at which two alternatives have equal annual costs. This may be done by calculating annual costs for each alternate, for different rates of interest, and interpolating.

One alternative is taken as the base, each of the others is compared with it, and the rate of return is computed. Those alternatives that do not show a rate of return that is sufficiently high are discarded. Economically speaking, the desirable alternatives are those that show a rate of return in excess of what is considered to be the minimum attractive return on investment. It is possible that some or all alternatives will be discarded as not yielding an attractive rate of return.

The rate of return method provides for selection of that project which will yield the most for the least expenditure of funds by giving priority to those projects yielding the highest rate. This is especially important in highway engineering where sufficient funds for proposed projects are not always available.

Benefit-Cost Ratio Method

The benefit-cost method is extensively used for highway, waterway, and other public works projects. This method involves the determination of the ratio of estimated benefits to estimated costs, and the ratio is used as the major criterion in determining whether or not the proposed expenditures should be recommended.

For each option, the annual road-user costs and the total annual cost for the highway are determined. The alternatives are then compared on the basis of benefit-cost ratios. The benefit-cost ratio is defined by:

$$\text{Benefit-cost ratio} = \frac{\text{benefits}}{\text{costs}} =$$

$$\frac{\text{difference in road user costs}}{\text{difference in highway costs}} = \frac{R - R_1}{H_1 - H}$$

Where:

R = the total annual road-user costs for the basic condition, usually the existing road

R_1 = the total annual road-user costs for the proposed improvement

H = the total annual highway costs for the basic condition, usually the existing road

H_1 = the total annual highway costs for the proposed improvement

The same formula is used to compute the benefit ratio for all alternatives, each compared with the same basic condition. On this basis, the relative values of the several ratios can be compared directly. A ratio greater than 1.0 indicates that the additional expenditure for the proposed improvement over the basic condition is justified. A ratio less than 1.0 indicates that the benefits are less than the costs, and in a road-user benefit sense the base condition is to be preferred over the proposed improvement. The alternative with the highest ratio is usually the most acceptable. To conclusively establish the most attractive alternative, however, it is necessary to compare all alternatives with each other, not just with the basic condition.

For comparable results it is essential that the same interest rates be used to evaluate the difference between annual cost of the various alternatives. The results obtained vary with the rate of interest assumed. In general, the interest rate only affects the denominator, and the higher the assumed interest rate the smaller the benefit-cost ratio.

ASSOCIATED ECONOMIC FACTORS

Sensitivity Aspects

Before an economy study can be undertaken, it is necessary in all methods of analysis, except the rate of return method, to adopt an interest rate. For all methods, including the rate of return, assumptions must be made for the useful life and the salvage value at the end of the useful life, for each component part of the highway improvement. In computing road-user costs for all methods, it is first necessary to estimate an ADT volume. All of these factors are based on assumptions that can affect the final result of an economic study, to a greater or lesser degree. Where important decisions are involved, a series of parallel solutions encompassing a reasonable range of assumptions may be warranted.

Effect of Interest Rates

The selection of an interest rate, or minimum attractive rate of return, significantly affects the result of an economic study. A decided advantage of the rate of return method lies in the fact that no predetermined rate of interest need be selected. All of the precautions that might be taken in carrying out an economic study are meaningless if the interest rate is inappropriate for the conditions under which the decision is made.

Low interest rates favor alternatives with large capital investments, and high interest rates favor alternatives with small capital investments. Low interest rates give more significant weight to happenings in the more distant future, where uncertainties of prediction are greatest, while high interest rates tend to discount the effect of future happenings.

Winfrey(2) states: "There is a pronounced tendency for those government officials and other persons who accept the principle of using a vestcharge rate for economic cost and analysis of economy to select a rate that is either the average interest rate the agency is paying on all outstanding debts or the rate at which bonds could be sold currently. Either of these rates, as such, is incorrect. They reflect the cost of financing or the cost of borrowing, and not the economic cost to the people who will furnish the taxes to service the debt. The bond interest rates do not reflect the worth of money or the worth of a proposed highway improvement to the individual citizens.

"A citizen who is paying 6 percent per annum on his home mortgage certainly would not consider it appropriate for a government to tax him for payment of a highway facility or water resource project which was economically justified by using interest rates of 3 percent per annum because that was the average rate on the particular government's bonded debt. It would be wonderful for the citizen if he could borrow money for personal use at a rate of 3 percent. He then could pay off his 6 percent mortgage, his 9 percent automobile loan, and his 18 percent reimproved facility may be considerably higher than traffic on the unimproved facility, and in many cases the projected ADT may be two or three times greater at the end of an analysis period of 20 or 30 years. Increased traffic volumes in the future may result in increased road-user costs due to congestion. Obviously, if the facility is operating at near capacity at the beginning of the analysis period, road-user costs can materially increase in the future if additional traffic causes greater congestion. The AASHTO procedure(4) of correlating vehicle running cost to type of highway operation is very desirable in this respect.

For important decisions, two analyses may be warranted; one based on a pessimistic estimate of growth, and the other on an optimistic estimate. With this approach the range of variation can be evaluated more accurately.

Determination of ADT for Period of Analysis

In order to evaluate road-user costs, it is necessary to determine the annual ADT for the period of analysis. AASHTO recommends three steps to determine the appropriate ADT:

- Estimate the ADT that will use the section upon its completion. This is the current traffic.

- Determine the number of years (usually taken as 20) for which the analysis is to be made, and the expansion factor for traffic on the section during this period. The traffic forecast period is independent of the average lives assumed for the various components that make up the capital investment.

- Calculate an expanded ADT that is a representative or an average value for the period of analysis. This is the ADT to be used.

It is not entirely clear as to the proper meaning of the third step. If the expanded ADT is used for the analysis, it would result in an overstatement of road-user costs. If the current ADT is used, it would result in an understatement of road-user costs. It is proposed that a representative average value be obtained by averaging the current ADT and the expanded ADT. But it is important to consider the appropriate ADT volume very carefully, as it has a significant influence on the analysis. This point is discussed further in the next section. It is necessary to separate traffic volume data by vehicle type, since different road-user costs must be used for each.

In comparing alternatives, the question often arises as to how to treat induced traffic, which might be a part of the total forecasted traffic growth for a proposed improvement, but would not necessarily be part of the traffic growth for the existing facility. "Induced traffic," as used here, includes all traffic that would not have occurred without the new facility. The magnitude of induced traffic is very much dependent upon the type and location of the proposed improvement. For example, induced traffic will normally be greatest for a limited-access facility connecting the central city to the suburban area. On the other hand, induced traffic will be of little significance within urban or rural areas.

There are different approaches to the treatment of the problem. These approches are best illustrated by the following example. Assume that the conditions on an existing road and a proposed new facility are:

Length of both alternatives = 10 miles

Unit road-user cost on existing = 12.5 cents per vehicle-mile

Unit road-user cost on proposed = 10.0 cents per vehicle-mile

Average ADT for analysis period for existing = 10,000 vehicles per day

Average ADT for analysis period for proposed = 15,000 vehicles per day

Annual cost of capital recovery for existing = 0

Annual cost of capital recovery for proposed (based on a capital investment of \$10 million (n = 30 years, i = 6 percent) = \$726,000

Annual road-user cost for existing = 10,000 x 10 x 0.125 x 365 = \$4,562,500

Annual road-user cost for proposed = 15,000 x 10 x 0.10 x 365 = \$5,475,000

The annual road-user cost for the proposed improvement exceeds the cost of the existing road. When considering the different ADT volumes anticipated for each facility, a benefit-cost analysis would not justify the construction of the new facility.

Analysis based on the ADT given for the existing:

Annual road-user cost for proposed =

$$10{,}000 \times 10 \times 0.10 \times 365 = \$3{,}650{,}000$$

$$\text{Benefit ratio} = \frac{912{,}500}{726{,}000} = 1.26$$

Analysis based on the ADT given for the proposed:

Annual road-user cost for existing =

$$15{,}000 \times 10 \times 0.125 \times 365 = \$6{,}843{,}700$$

$$\text{Benefit ratio} = \frac{1{,}368{,}700}{726{,}000} = 1.88$$

Analysis based on total benefits to be derived from proposed:

Annual total benefits =

$$(0.125 - 0.10) \times 15{,}000 \times 10 \times 365 = \$1{,}368{,}700$$

This analysis leads to the same result as the analysis based on the ADT given for the proposed.

In the above example, when the same ADT is used in comparing alternatives, construction of the proposed facility would be justified.

Admittedly, there is no single answer to cover all situations, but the most reliable method is to base the comparison on the ADT given for the proposed improvement. The annual cost of capital recovery for the proposed improvement is significant in the comparison, and it is based on the capital investment required to provide a certain level of service for the traffic to be accommodated by the new facility. Therefore, the ADT on the proposed improvement is the more meaningful one to use in the analysis.

The rules for selecting projects and establishing priorities are given in the next section, which was excerpted from the AASHTO "Manual on User Benefit Analysis."

Compound Interest Factors

The economic techniques for evaluating options use various compound interest factors; the formulas for these are contained in table 13.2.

TABLE 13.2. COMPOUND INTEREST FACTORS

Factor	Use	Formula or Relationship
Present worth of a single sum (PW)	To find P, given F	$\frac{1}{(1+i)^n}$
Uniform series present worth (SPW)	To find P, given A	$\frac{(1+i)^n - 1}{i(1+i)^n}$
Compound amount of a single sum (CA)	To find F, given P	1/PW
Capital recovery (CR)	To find A, given P	1/SPW
Sinking fund (SF)	To find A, given F	PW/SPW

Where:

P = a present sum of money or equivalent value

F = a future sum or equivalent future value

A = uniform annual end-of-period payments

n = number of interest periods, usually in years

i = interest rate per period

DECISION RULES FOR SELECTING OPTIONS

The accepted decision rule for economic efficiency is to select the set of improvement projects or project alternatives that yield the greatest net present value (NPV). Procedures for selecting projects and project alternatives in order to maximize NPV vary according to the presence and nature of budget limitations or constraints, the independence or mutual exclusivity of projects or alternatives, and the presence or absence of investment costs. In all cases, the optimum set or portfolio of projects can be selected through a linear or dynamic program that takes any budget constraints into account. In effect, this procedure searches all combinations of projects for the set that maximizes NPV under given constraints.

This search can also be done manually for a small number of projects, but a manual search of all combinations becomes cumbersome for a large number of projects. Instead, simpler sequential procedures can be used for approximate (though usually accurate) project selection and prioritization purposes, if care is used in distinguishing between three different cases. These manual procedures are based on benefit/cost ratios that, when properly applied, can provide an indication of the relative economic efficiency of project alternatives when budgets are constrained. For ease of exposition, the simplest case—with no budgetary constraints—is considered first, followed by discussions of single and multiple transportation budget constraints.

Case 1

In the simplest situation where there are no budgetary constraints, or where individual projects are being considered in isolation, all independent projects (those whose selection would not preclude selection of other projects) with positive NPV's should be selected. This is equivalent to accepting all independent projects with benefit-cost ratios above 1.0 according to either equation (1) or (2) as defined below. In the case of mutually exclusive alternatives (e.g., a low-cost and high-cost alternative to the same type of improvement at the same site) when there are no budgetary constraints, the alternative having the highest positive NPV can be selected. Equivalently, mutually exclusive alternatives can be selected by increments, provided that each increment of cost results in a positive increment of NPV.

In using benefit-cost ratios for determining which mutually exclusive alternative to adopt, it is not enough to compare each alternative with the do-nothing alternative. To ensure maximization of net present value, selection must be made in increments, with each increment of cost justified by testing whether the associated incremental benefit-cost ratio is greater than 1.0.

The following procedure is recommended for using benefit-cost ratios for choosing the most economically desirable alternative from among mutually exclusive alternatives when no budgetary constraints pertain.

1. Array all mutually exclusive alternatives in ascending order of cost relative to the do-nothing alternative.

2. Using benefits and costs relative to the do-nothing alternative, compute the ratios of the present value of benefits to present value of costs for each alternative. Tentatively accept the lowest-cost alternative with a ratio greater than 1.0. If no such alternative is found, the procedure terminates, and the do-nothing alternative is selected. Otherwise, proceed to step 3.

3. Consider the next-higher-cost alternative (relative to the tentatively accepted one) and compute the ratio of the present value of incremental benefits to the present value of incremental costs, where incremental benefits and costs are those in excess of the benefits and costs associated with the tentatively accepted lower-cost alternative. A ratio greater than 1.0 implies that the incremental benefits are greater than and therefore justify the incremental costs. If the ratio is greater than 1.0, discard the previously accepted alternative and tentatively accept this alternative. This alternative becomes the basis for comparision with the next-higher-cost alternative. If the incremental benefit-cost ratio is not greater than 1.0, the previously accepted alternative remains and the next-higher-cost alternative is judged in terms of an incremental benefit-cost ratio relative to the previously accepted alternative.

4. Repeat step 3 for all higher-cost alternatives. Select as the economically most-justified alternative, the highest-cost alternative with an appropriately calculated incremental benefit-cost ratio greater than 1.0.

A numerical example of incremental benefit-cost analysis for a pair of mutually exclusive projects is presented later.

Case 2

In the situation where there is a budget constraint on highway or transit investment, the decision rule changes to maximizing NPV for the available budget. This can be accomplished by selecting the combination of projects that have maximum present value but, in total, do not violate the budgetary constraint. For large numbers of projects this can be accomplished by linear or dynamic programming procedures. Equivalently, if there are several independent projects to choose from, selection of projects in order of declining benefit-cost ratios will maximize NPV for a given investment budget. For this purpose, benefit-cost ratios can be defined with maintenance and operating (M&O) costs and residual value in the numerator, as follows:

$$BC = \frac{PV(\Delta U) - PV(\Delta M) + PV(\Delta R)}{PV(\Delta I)} \quad (1)$$

Where:

PV = the present value of the associated parenthetical amount or series of amounts over time, discounted at the selected discount rate

ΔU = reduction in the series of annual highway or transit-user costs due to the investment (costs without the improvement less costs with the improvement)

ΔI = increased investment costs due to the project

ΔR = increase in residual value due to the project at the end of the life of the project

ΔM = increase in the series of annual maintenance, operating, and administrative costs, or M&O costs for short, due to the investment (costs with the improvement less costs without the improvement); if these costs decrease, they are then negative costs and are, in effect, added to user benefits

The rationale for deducting M&O costs from benefits in equation (1) follows from the stated objective of maximizing the present value of the excess of discounted benefits over discounted costs per unit of highway or bus transit investment. Residual value belongs in the numerator in this case because it can be considered equivalent to a positive future cash flow or benefit.

To use equation (1) for ranking independent (nonmutually exclusive) improvements within a given construction or investment budget, the project with the highest ratio is selected first and other projects are added to the list until the budget is exhausted. If the last project selected exceeds the budget, it can be either—in the usual order of economic attractiveness: (1) started during the budget year and completed out of the next year's budget; (2) displaced to a later year by a lower-cost project that can fit within the budget; or (3) postponed without replacement. In a linear programming approach, such alternatives should be included in the initial problem formulation.

Non-independent projects can be selected in the same fashion so long as: (1) incremental benefit-cost ratios are used for evaluating each increment of cost separately, starting at the lower cost alternative; and (2) lower-cost projects are displaced from the accepted list when a mutually exclusive higher-cost alternative is selected.

Case 3

In the more common situation where both the investment budget and future M&O budgets are constrained, equation (1) is not appropriate, and a linear or dynamic programming solution is preferred. However, the following version of the benefit-cost ratio, with all government agency costs in the denominator, will serve for approximate identification of the order of project desirability when funds are interchangeable between investment and operating budgets (i.e., the budget constraint is joint rather than separate):

$$B/C = \frac{PV(\Delta U)}{PV(\Delta I) + PV(\Delta M) - PV(\Delta R)} \quad (2)$$

This version of the benefit-cost ratio answers the question: what is the return per dollar on the total government cost of a highway or bus transit improvement? However, the set of projects selected in the order of declining benefit-cost ratios of this type will not always maximize the net present value of future cash flows for a given investment budget, because more than investment costs are included in the denominator of the ratio.

Discussion

In the first of these three cases, it does not matter whether equation (1) or (2) is used because they are equivalent when the question is only whether a benefit-cost ratio is 1.0 or greater. As noted by Fleischer(5), a ratio cannot be changed from less to more than unity, or vice versa, by shifting a quantity between the numerator and the denominator. Letting that quantity be K in this case:

$$\text{if} \quad \frac{B - K}{C} > 1$$

$$\text{then} \quad B - K > C$$

$$B > \left[C + K\right]$$

$$\text{and} \quad \frac{B}{C + K} > 1$$

However, shifting K between the numerator and denominator may affect the relative ranking of two projects by the benefit-cost ratio, as will be illustrated in the example that follows shortly.

Case 2 is appropriate for typical investments where the benefits consist of a stream of revenues that can be used to meet any increased maintenance or operating costs. This may appear to include bus transit improvements that result in added fare revenues, but it does not do so well because the added fares will not usually be sufficient under today's conditions to offset increased operating costs. Case 2 also does not represent highway projects well because the user benefits from highway investments cannot be respent by the highway agency. Moreover, highway M&O budgets are constrained by the difficulty and infrequency with which highway user charges are raised. If future highway user charges to transit revenues can be raised at will to take advantage of attractive projects that would nevertheless increase the future need for revenues beyond prospective budgets, then case 2 applies.

Case 3 is closer to both highway and transit budget realities, where the maintenance and operating budgets are also constrained and hence relevant to the choice of projects. Also, the benefit-cost ratio for case 3—(equation (2))—avoids the occasional problem of very high or infinite benefit-cost ratios when relevant investment costs are or approach zero (such as might occur in a bus transit project). Hence, equation (2) is recommended for all highway and transit applications, but only as an approximate indicator of project desirability. The ultimate test should be whether the set of projects selected maximized NPV within applicable budget constraints.

Illustration

To illustrate the foregoing cases with an example, assume that projects X, Y, and Z are being considered for funding and that project Y can be accomplished by either of two mutually exclusive alternatives, Y1 and Y2. The relevant data and indexes are presented in table 13.3; residual value is ignored for simplicity. All investment costs are assumed to accrue at the present time and to be subject to investment budget constraints.

These data are considered below as if the budget conditions for the analysis correspond in turn to each of the foregoing cases. Case 3 becomes fairly complex because of the dual budget constraints, and hence suggests the advisability of using a linear programming solution. Both cases 2 and 3 illustrate the handling of the marginal project, i.e., that project which if ranked against other projects would just meet or exceed the budget. The examples for cases 2 and 3 produce the same results: initiating a marginal project that exceeds the budget constraint (and finishing it at a later period) is preferable to postponing the entire project or substituting different but lower-priority projects that come within the budget constraint.

Case 1 Example

If there is no budget constraint or if these three projects are being considered in isolation, the projects X and Z and project alternative Y2 should be selected because they all have positive net present values and for project Y, alternative Y2 has the highest NPV. If benefit-cost ratios (B/C) are used, the project selection process would proceed as follows:

TABLE 13.3. PRESENT VALUES AND RATIOS FOR PROJECTS X, Y, AND Z

	X	Y1	Y2	Y2-Y1	Z
Present value of user benefits, PV(ΔU)	100	98	128	30	26
Present value of maintenance and operating cost increases, PV(ΔM)	10	2	22	20	20
Present value of investment cost, PV(ΔI)	10	12	13	1	0
Benefit-cost ratio, equation (1)	9	8	(8.2)	10	--
Benefit-cost ratio, equation (2)	5	7	(3.7)	1.4	1.3
Net present value (NPV)	80	84	93	9	6

Note: For numbers in parentheses, see text for relevance to project selection decision.

- Select project X and Z, and tentatively select Y1, because their B/C ratios are greater than 1.0.

- Test alternative Y2 by either B/C ratio for the increment Y2 - Y1; since they are greater than 1.0, select Y2 and displace Y1, the mutually exclusive alternative from the list of selected projects. Total net present value would be 80 + 93 + 6, or 179 units.

The relevant ratios for considering Y2 are based on the comparison Y2 - Y1 because Y1 (compared with doing nothing) has a ratio over 1.0. If Y1 had a ratio below 1.0—signifying negative NPV—it should be rejected, and Y2 would then be compared with doing nothing by the ratios shown in brackets.

Case 2 Example

Assuming an investment budget of 22.5 units, the selection process could proceed as follows:

- Select project Z, which has no initial cost and positive NPV, hence a B/C ratio of ∞ by equation (1).

- Select project X, which has the next highest benefit-cost ratio according to equation (1) (except for Y2 versus Y1, which is not relevant until Y1 has been tentatively selected).

- Tentatively select Y1, with a ratio of 8 according to equation (1). The total investment cost so far is 22 units.

- Test Y2 by the same equation. A ratio of 10 for Y2 - Y1 indicates the acceptability of the project on economic grounds, but the added initial cost of 1 unit causes the budget of 22.5 units to be exceeded by one-half unit. Hence, a decision is needed whether to (1) begin Y2 instead of Y1 and complete the project with funds from next year's investment budget; (2) do projects Z and X, postpone a decision between Y1 and Y2 to next year, and look for another project or projects with a B/C ratio of 1.0 or greater, and investment costs of up to 12.5 units to fit within this year's investment budget; or (3) do projects Z, X, and Y1, and look for another attractive project or projects with investment costs up to 0.5 units to fit the investment budget. On economic grounds, the alternative from these three that results in greatest NPV should be selected. In the example, it would be the first alternative, beginning Y2, assuming that Y2 has a higher B/C ratio than other projects that might better fit the available budget.

Note that project Y2, though it has an incremental ratio of 10 compared with Y1, is still not preferred to project X because the total ratio of Y2 compared with doing nothing is only 8.1 compared with 9 for project X. It is easy to believe that the higher NPV of alternatives Y2 and Y1 compared with X indicates their economic superiority over X. But this is true only in case 1, the absence of budget constraints.

Case 3 Example

Again assume an investment budget of 22.5 units and add the constraint of an M&O budget for this year not yet committed to other purposes of 2.05 units plus an added 1.5 units (3.55 in total) by the following year because of increased operating revenues or decreased maintenance costs. It is first necessary to derive the annual M&O costs for each alternative from the equivalent present value of M&O costs. Assuming a study period of 25 years, and a discount rate of 4 percent, the capital recovery factor (that can be calculated from the formula given earlier) is 0.0640 and the annual M&O costs for each project or alternative project are estimated as follows:

X:	0.64
Y1:	0.13
Y2:	1.41
Z:	1.28

The selection process for this case utilizes the second B/C ratio, equation (2). Project Y1 ranks ahead of project X on this basis (a ratio of 7, rather than 5), reversing their order of preference according to equation (1), because Y1 has a low proportion of future maintenance costs in relation to initial costs. The selection process is as follows:

1. Tentatively select Y1 for the reason given.

2. Consider Y2 in comparison with X and Z. Since X has a ratio of 5 compared with 1.4 for Y2 - Y1 and 1.3 for Z, X is selected next.

3. Project Y2 is next in order of preference, but because it would cause the investment budget to be exceeded by 0.5 units, the same type of alternatives considered for Y2 in the preceding case should be evaluated:

   A. Initiate Y2 in the budget year and complete it out of next year's budget.

   B. Substitute another project for both Y1 and Y2, so Y2 can be carried out the following year (in the preceding alternative, doing Y1 would preclude ever doing Y2).

C. Substitute a project with a lower initial cost for Y2.

<u>Alternative A</u> – Consider the first of these alternatives, the initiation of Y2 this year and completing it out of next year's budget. Budget information and relevant NPV's are summarized in table 13.4. Y2, which has replaced Y1, is started in the current year at an investment cost of 12.5. Completion of

TABLE 13.4. INITIATE Y2 AND COMPLETE NEXT YEAR
(Budget and NPV Information)

Choice	Investment Cost	M&O Cost	NPV
Budget Year			
X	10.0	0.64	80.0
Y2 (Start)	12.5	--	--
	22.5	0.64	80.0
Next Year			
Y2 (Completed)	0.5	1.41	91.4
Z	0.0	1.28	5.8
Two-Year Total	0.5	3.33	177.2

the project will be accomplished in the following year with the expenditure of 0.5 of investment. Because of the stretched-out completion, it is assumed that the realization of benefits accruing from the project would be deferred by an average of one-half year over the life of the project and that the increase in M&O costs would not occur until the next year. Assuming a 4 percent discount rate—or an approximate 2 percent discount rate for one-half year— deferring of benefits for one-half year results in a revised (from 128) present value of benefits of 125.5 (128 – 1.02); the year's deferral of M&O cost increases and of the investment cost of 0.5 reduces their (negative) contribution to NPV from a total of 22.5 to 21.6 (22.5 – 1.04). Revised NPV for Y2 is thus 125.5 - 12.5 - 21.6, which equals 91.4. Deferral of all of project Z until next year results in a revised NPV of 5.8 (6 – 1.04).

Note that the budget year investment budget is exhausted, but there is still 1.41 units left in the budget year M&O budget of 2.05. Addition of Y2 and Z next year will increase annual M&O costs to 3.33 units, which is within the assumed M&O budget of 3.55 for that year. Complete evaluation of this possibility would require information regarding possible additional projects for the "next year."

Alternative B – The second possibility, substituting Z for Y1 and Y2 in the budget year and postponing Y2 until next year, with the NPV of Y2 discounted by 4 percent to reflect the one-year deferment, gives the results shown in table 13.5.

TABLE 13.5. SUBSTITUTE ANOTHER PROJECT

Choice	Investment Cost	M&O Cost	NPV
Budget year			
X	10	0.64	80.0
Z	0	1.28	6.0
	10	1.92	86.0
Next Year			
Y2	13	1.41	89.4
Two-Year Total		3.33	175.4

Alternative C – Now consider the results if project Z is chosen in place of Y2, but Y1 is carried out, to exactly exhaust the M&O budget. Table 13.6 shows the results for this case.

The NPV for solutions (A) and (B) are very close, so a choice between those solutions could be made on other grounds. The NPV for solution (C) is lower, essentially because it forces a choice of Y1, precluding Y2 from ever being carried out although it is an economically attractive increment of investment.

TABLE 13.6. SUBSTITUTE A PROJECT WITH A LOWER INITIAL COST

Choice	Investment Cost	M&O Cost	NPV
Y1	12	0.13	84
X	10	0.64	80
Z	0	1.28	6
	22	2.05	170

REFERENCES

1. Pignataro, Louis, J. "Traffic Engineering Theory and Practice."

2. Winfrey, Robley. "Economic Analysis for Highways." International. 1969.

3. Glancy, David M. "Utilization of Economic Analysis." State Highway Dept. HRR 77.

4. AASHTO. "Manual on User Benefit Analysis." 1977.

5. Fleischer, E. A. "Numerator-Denominator Issue in the Calculation of Benefit-Cost Ratios." Transportation Costs. Highway Research Record No. 383. TRB.1972.

The use of simulation models for diagnosis and prediction is described in chapter 5 of this volume, which outlines the range of simulation models and their capabilities. During those stages of implementing a system where the designer wishes to evaluate a series of choices between likely alternatives, these models can also be used for design and evaluation.

For example, the CORQ model can provide detailed treatments for major intersections. The FREQ6PE model is useful when priority entry control is being evaluated. When designing incident control strategies, the INTRAS model may be used to evaluate which strategy is the most effective.

During the design stage of a project, the engineer should consider using the models described in chapter 5 in conjunction with the cost-effectiveness evaluation techniques outlined in chapter 13. This will provide a quantification of the likely results of implementing a particular option without undergoing the expense of construction and testing.

SECTION GUIDE

IMPLEMENTING SOLUTIONS

Three key elements in the successful implementation of a freeway management system that are discussed in the remaining chapters are:

- Financial planning (chapter 15)
- Staged construction (chapter 15)
- Interagency cooperation (chapter 16)

The Federal Highway Administration can provide funds for the initial capital costs of freeway management projects. Long-term operating costs are normally borne by the states and concerned jurisdictions. The financial planning associated with this balance is critically important.

The staged construction of a freeway management system allows these financial expenditures to be spread out over time, rather than severely impacting the budget of a particular fiscal year.

As previously discussed, there are many public agencies and private groups that are involved in making the freeway management system successful. Although the police and highway departments are the most visible of these agencies, others must also be consulted in the planning and operation of a facility.

The success of any project, however, will ultimately depend on the careful organization and cooperation of those involved. To assure interagency cooperation, it is necessary to promote good communication. Frequent review meetings among agencies and individuals involved in the design and operation of a freeway management project, will aid in recognizing problems and resolving organizational conflicts.

After the ideal components of a freeway management system have been determined, an appropriate implementation plan must be developed. It is likely that the agencies involved in funding the project will not be willing or able to commit the financial resources necessary for the construction of all system components at the same time. However, this should not be considered as an overwhelming obstacle. Many freeway management systems have developed from small beginnings, and a staged approach to implementation can have several advantages.

A staged implementation plan affords time to give all actions thorough consideration, to reach agreements with cooperating agencies, and to thoroughly prepare detailed freeway control designs and construction cost estimates. Long lead times avoid the last minute rush to get the plans and specifications finished ahead of the letting dates.

A staged implementation program can be continually updated by project priority selections and can provide a financial base for keeping the whole project on an even keel. All units of the department may then work on their own toward specific common goals. Employment levels become more uniform, the necessary skills can be made available, and each implementation project can be coordinated with others as desirable for engineering consistency, production efficiency, and budgetary constraints.

FINANCIAL PLANNING

The individual responsible for developing an implementation plan must keep in mind the financial and political realities of the project environment. It may be necessary to develop a series of alternatives based on different funding levels. It would also be prudent to start with an element of the program that was visible and could produce some immediate improvements (e.g., ramp metering or better onsite incident management) to generate community and political support for continued system expansion.

The various funding sources, and the areas in which these funds can be applied, must also be carefully considered. Funds from the Federal Highway Administration are often available for capital construction. State and local funds are normally used for operational expenses. Because of this dichotomy an alternative that has a high capital construction cost and low operational cost may be preferable to an alternative that has a lower capital cost and a higher operational cost.

PROJECT MANAGEMENT

The control of costs and schedules is the most important aspect of managing the implementation of a project. Although costs and schedules can be monitored manually, the complexity of many projects often requires the assistance of a computerized management system. Several existing systems offer the following features to assist the project manager:

- Project Planning and Scheduling — Including Critical Path Method network planning with up to 32,000 activities. The outputs of this activity can include schedules, bar charts, and CPM charts aggregated to any level necessary for project management or management review.

- Resource Allocation and Scheduling – Offers the capability for assigning contractors, personnel, and materials resources by time period to specific project activities. It includes the ability to schedule and change priorities to adjust schedules based on project delays, to assign resources individually, or to aggregate them.

- Cost Estimating – Should be linked to other elements including the resource, project planning, and materials management data bases to provide the ability to maintain a detailed cost model of the project. The cost estimating data base may accept vendor pricing data, labor and equipment rate tables, crew and equipment mix tables, construction cost tables, escalation indexes and productivity factors. The ability to relate the cost model to other elements of the data base will permit continuous tracking of project costs and estimation of cost to complete.

- Cost Management – Includes elements of the cost estimation process and provides calculation of cash flow and financial accounting. It provides outputs summarized by cost code, work package code, or any other coding scheme that may be desired. It may include escalation rates and contingency allowances as part of the cost management process. It should permit examination of alternative design and schedule scenarios and should be capable of analyzing various payment periods.

- Material Management – Can be integrated into both the planning and cost control of the project. For example, it is possible to link an equipment test to the purchase order and shipping registers and have this information related to the project schedule. It can also be used to provide data to the cost model by providing bulk purchase discounts, vendor progress payment schedules, delivery acceleration payments, and material cost escalation. It also has the capability to track material shipment, site receipt, storage, and issue for construction. The ability to track all material from order to issue for construciton should significantly improve the ability to control inventory theft by tracking all material used for the system.

- Records Management – Can provide a valuable cross reference between project drawings and other data items. In addition to performing library functions, this feature can have the capability of providing summaries of the status of plan development and preparation of as-built drawings, as well as other project documents.

STAGED CONSTRUCTION

The illustrative example given below outlines how a medium-sized freeway traffic management system can be built in stages to eventually form the final design. Imagine a freeway system that has congestion problems resulting from excessive demand and incidents. Following selection of the options available and evaluation of their effects, selection of the most economic candidates, and an agreement to cooperate from the participating agencies, the following freeway management requirements can be identified:

- Fifty miles of the freeway should have electronic surveillance and ramp control.

- A 10-mile section in the CBD needs 12 closed-circuit television (CCTV) cameras.

- The existing police patrols should be equipped to deal with minor incidents.
- Additional highway patrols will be provided; these will be operated by the highway agency.
- Forty-five ramps will have traffic-responsive control.
- CB radio monitoring will be used in the central control and agency vehicles, but not in the police vehicles.

Year 1 – Equip police vehicles (providing immediate saving at a minimum cost), write specifications and contract for cable ducting for future ramp control and CCTV. Select site for central control room. Install surveillance loops in CBD area.

Year 2 – Order patrol vehicles and CB radio equipment. Award cabling contract, begin installation. Continue loop installation and add metering at the five most critical ramps. Acquire land for central control room and begin construction.

Year 3 – Install metering equipment at an additional eight ramps and continue loop installations. Start service patrol and CB monitoring of incidents. Continue cable ducting installation and control room construction. Write specifications for central control computer and its peripherals.

Year 4 – Install more loops and meter 10 more ramps. Expand service patrol. Finish control room construction, continue cable and ducting installation. Order computer and peripherals, write specifications for CCTV and data transmission equipment.

Year 5 – Install ramp metering equipment at an additional 11 ramps and expand loop installation. Take delivery of and test computer, its peripherals, and data transmission equipment. Finish cable ducting. Test CCTV equipment.

Year 6 – Install metering at the final 11 ramps and CCTV equipment. Change ramp control to traffic-responsive mode –turn computer on.

This hypothetical example is obviously oversimplified, but it does illustrate the staged construction approach. In the first year there will be small benefits from the equipping of police vehicles. In the second year, more reductions in delay can be made with the installation of ramp metering at the first ramps. As more ramps are included in the systems, the benefits increase until the sixth year when the entire planned system is operational. Table 15-1 is a listing of the installation contracts for the Chicago Area Expressway Surveillance and Control System(1). This system, which includes several freeways, has been growing and evolving over a 15- year period.

TABLE 15-1. CHICAGO AREA EXPRESSWAY SURVEILLANCE AND CONTROL INSTALLATION CONTRACTS

Award Date	Locations	Major Work Items	Cost
January 1967	I-94, Dan Ryan	SI, SE, RM	$ 125,700*
August 1967	I-290, Eisenhower	SI, RM	171,400
August 1967	I-290, Eisenhower	SE, CC, RD	169,800
August 1969	I-94, Edens, Kennedy	SI, RM	650,000
August 1969	I-94, Edens, Kennedy	SE, RD, DS	347,300
October 1969	I-90, I-94, Dan Ryan	SI	422,500
November 1969	I-90, I-94, Dan Ryan	SE, RD	166,100
November 1969	Surveillance Center	CCS	378,100#
March 1970	I-55, Stevenson	SE, RD	165,000
April 1970	I-55, Stevenson	SI	479,700
June 1971	I-90, Kennedy	SI, SE	137,900*
September 1971	I-290 Eisenhower	SI, SE, RM	368,300
September 1972	I-94, Calumet, Kingery	SI, SE, RM, AD	666,600**
July 1973	I-290, Eisenhower Extension	SI, SE, AD	657,100
October 1975	I-57, I-94	SI, SE, RM, DS	875,100#
April 1978	I-55, I-90, I-94	SI, SE, RM	878,700#
February 1979	I-90, I-94, Dan Ryan	CMS	199,900#
March 1979	I-94, Edens	SI, SE, RM	450,000***
November 1980	I-290, Eisenhower	SCC	2,043,500
May 1982	Traffic Systems Center (SCC)	CCS	1,368,100#
May 1982	Traffic Systems Center (SCC)	D/W	82,000
			$ 10,802,800

Notes: 1. Excludes initial Eisenhower Expressway research installation contracts (between 1961 and 1964), as well as any maintenance or rehabilitation contracts.

2. All funding with Federal participation, primarily as interstate construction, except where noted (*) as "state-only" funds.

3. # Includes initial maintenance period, not Federally funded.
** Includes Federal participation in leased data line costs.
*** Estimated cost.

Legend: Major Work Items

SI Surveillance field installations
SE Surveillance equipment installations
RM Ramp metering installations
CC Ramp metering control consoles
RD Route map displays
DS Map display scanners
CCS Central computer system
AD Area map displays
CMS Changeable message sign system
SCC Surveillance/control center building
D/W Dewiring/wiring equipment

REFERENCES

1. McDermott, J. M., Director, Traffic Systems Center, Illinois Department of Transportation—private communication. February 1983.

CHAPTER 16. INTERAGENCY COOPERATION

An important consideration in the implementation and operation of a freeway management system is the cooperation of the agencies responsible for providing the needed response. Frequently, more than one department, agency, or jurisdiction is involved. Because the priorities within each agency are often different, it can be difficult to achieve full cooperation. To overcome these differences it is sometimes necessary to create a team composed of representatives of the major agencies and/or governmental entities. Responsibility for initiating and coordinating all incident management activities is given to this team.

IMPLEMENTATION

The cooperation of the various participating agencies during system implementation can be a determining factor in the success or failure of a freeway management system. Periodic meetings of these agencies during the detailed planning and construction phases can help identify and resolve construction issues while fostering the interagency rapport that is essential for project success.

These interagency meetings will provide the lead agency having primary responsibility for construction (probably the highway department) with insight into the operational problems that may be handled by another agency (perhaps the police department). For example, the system design may include provisions for HOV lanes. The overall design of this facility would come from the department of transportation. Details on how the HOV restriction would be enforced and the required frequency and size of refuge areas for ticketing violators should be closely coordinated with the police department. The synergism resulting from this cooperation will produce a better working facility than could be produced if the design had been developed in isolation.

These meetings will also provide an opportunity for the people from the highway and police departments, who will be operating the facility, to get to know one another. Care must be exercised, however, to ensure that these meetings do not become confrontations between these individuals and the agencies that they represent.

OPERATION

There are six basic organizational concepts that can be used for incident management.

Police Department

In most cities, the state or local police agency has operational responsibility for the freeway system. This duty evolves quite naturally from traditional police concerns of traffic law enforcement, traffic control, and accident investigation. Some departments view it as control over the vehicles in the freeway environment, in contrast to highway department responsibility for the stationary elements. Given this high degree of involvement in freeway operations, many police agencies regard incident management as a logical extension of their existing responsibilities. Hence, police agencies tend to implement incident management techniques that build on their basic patrol functions.

A meaningful commitment, however, requires the assumption of a service role that may not be practical for some police agencies, due to limited experience or inadequate resources. Under such circumstances, incident management responsibility may fall on the next most logical public agency, a highway department or department of transportation (DOT).

Police and Highway Department

The traditional construction and maintenance responsibilities of highway departments and DOT's have necessitated the acquisition of many of the vehicles and much of the equipment used for the cleanup and removal of freeway incidents. Maintenance and operational responsibilities may be divided among several jurisdictions. Furthermore, some DOT's have acquired an additional incident management role as a result of operating electronic surveillance and control systems. These activities provide a sound base upon which to build a more comprehensive incident management system under DOT control. Many of the potentially new incident management techniques, however, involve responsibilities that are usually those of police agencies. A DOT-sponsored incident management system can only be seriously considered with the approval and cooperation of the local police agency.

Highway Department

A DOT-sponsored incident management system can operate effectively under two different organizational configurations: as the joint responsibility of the DOT and the police department, with each agency still performing its traditional role but having closer coordination with the other; or as the sole responsibility of the DOT, with the agency assuming what were previously police incident management responsibilities, such as providing aid to motorists or monitoring CB radio reports.

Point Facility Authorities

For bottleneck or point facilities such as bridges and tunnels, the operating authority is often the initiator of an incident management system. These facilities present special problems, since shoulders or emergency service lanes are usually nonexistent and diversion opportunities limited. Consequently, measures to reduce delay and congestion are of great importance to the authority, which, because of its operational experience, is the most logical group to oversee implementation. Furthermore, police agencies and highway departments generally play only supporting roles with respect to incident management on point facilities. Point facility incident management systems tend to emphasize their surveillance measures since detection over a short stretch of highway can be virtually instantaneous without being prohibitively expensive. This fact explains the low-cost classification of options using closed-circuit television and loop detectors for point facilities, even though their installation throughout an entire freeway system would require a large investment.

Citizens Groups

A citizen group may take the lead in implementing incident management options, particularly when it perceives official activities to be inadequate. Because of citizen groups' volunteer nature and limited resources, these incident management options tend to be labor-intensive and often consist of methods for improving the speed and quality of the incident-related information transmitted to public agencies and the media. Citizen involvement frequently develops out of existing organizations such as citizens band (CB) radio clubs and civic associations. It can then be easily broadened to include crime-reporting activities for the benefit of local law enforcement agencies.

Incident Management Team

The final incident management system, the freeway incident management team or authority, is considered to be the most effective due to the consolidation of all response and surveillance responsibilities into one administrative unit. This allows for the complete management of incidents by a single multidisciplinary team composed of police, highway, and other public agency personnel, with possible assistance by citizen volunteers. Such an authority is usually not available, so a team system is likely to require an intense, well planned organization and administrative effort at its inception. The potential benefit (i.e., elimination of the coordination and communication problems that reduce the effectiveness of other systems), however, makes the extra effort worthwhile. An additional advantage of the team system is that all incident management options are available to it, since it represents a combination of the other five systems. This integrated freeway management team composed of police, highway, and other public agency personnel is considered to be the most effective form of staffing. The incident management team assures a unified effort in operating, maintaining, and updating the system, and provides the opportunity for developing a skilled staff that can be increasingly responsive to system requirements.

Figure 16.1 reviews the options available for incident management. This figure shows which agencies generally have the operational responsibilities for each of the options. These divisions are by no means rigid, as in specific locations one agency often takes on the responsibility for many of the options listed. In the case of point facilities at bridges or tunnels, the operating agency can be considered to perform similar functions to that of a highway department.

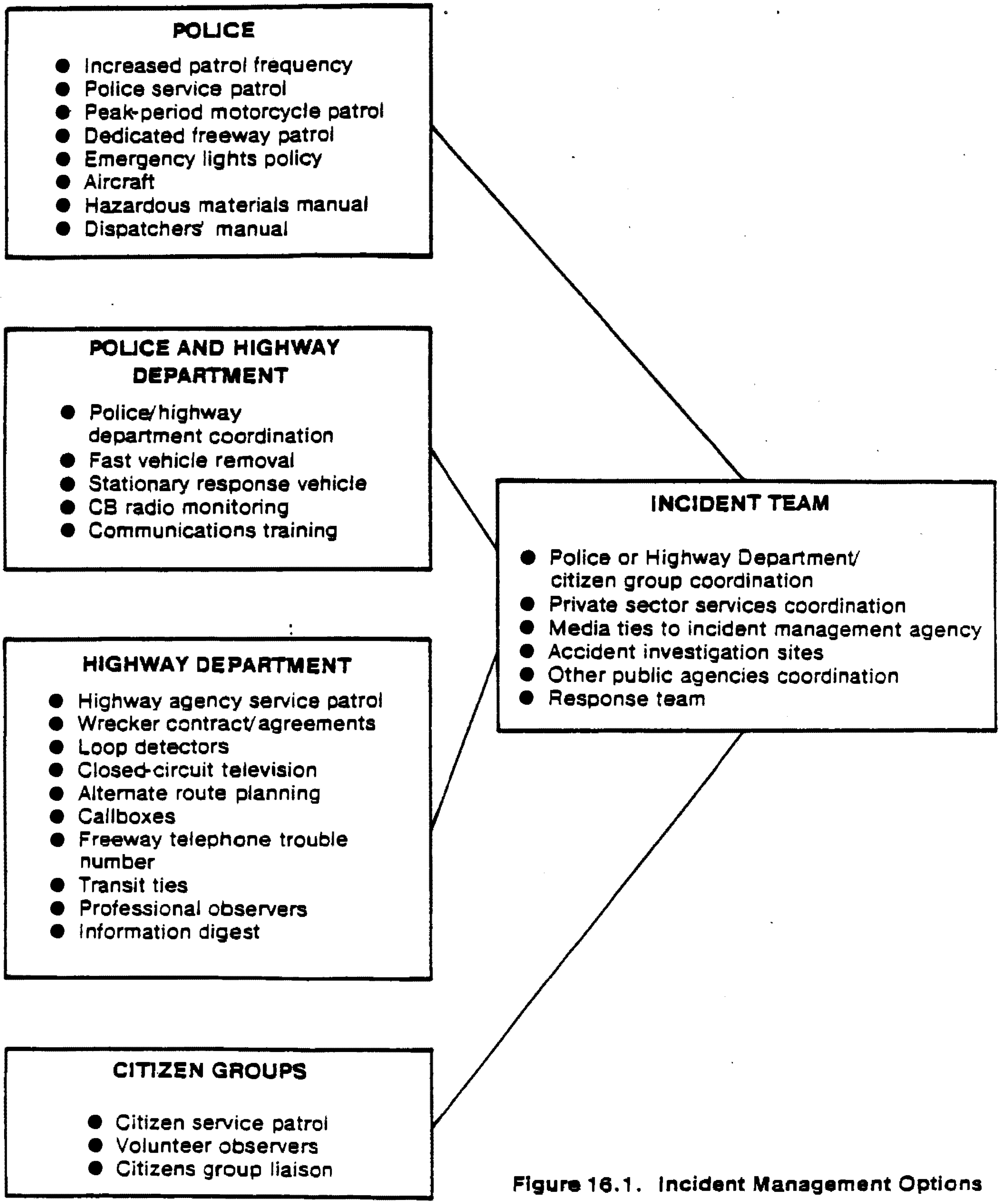


Figure 16.1. Incident Management Options