TRANSPORTATION RESEARCH RECORD

533

Freeway Entry, Flow, and Control

10 reports prepared for the 54th Annual Meeting of the Transportation Research Board

TRANSPORTATION RESEARCH BOARD

NATIONAL RESEARCH COUNCIL

Washington, D. C., 1975

Transportation Research Record 533
Price $5.60
Edited for TRB by Marianne Cox Wilburn

Subject areas
53 traffic control and operations
54 traffic flow

Transportation Research Board publications are available by ordering directly from the Board. They are also obtainable on a regular basis through organizational or individual supporting membership in the Board; members or library subscribers are eligible for substantial discounts. For further information, write to the Transportation Research Board, National Academy of Sciences, 2101 Constitution Avenue, N.W., Washington, D.C. 20418.

These papers report research work of the authors that was done at institutions named by the authors. The papers were offered to the Transportation Research Board of the National Research Council for publication and are published here in the interest of the dissemination of information from research, one of the major functions of the Transportation Research Board.

Before publication, each paper was reviewed by members of the TRB committee named as its sponsor and accepted as objective, useful, and suitable for publication by the National Research Council. The members of the review committee were chosen for recognized scholarly competence and with due consideration for the balance of disciplines appropriate to the subject concerned.

Responsibility for the publication of these reports rests with the sponsoring committee. However, the opinions and conclusions expressed in the reports are those of the individual authors and not necessarily those of the sponsoring committee, the Transportation Research Board, or the National Research Council.

Each report is reviewed and processed according to the procedures established and monitored by the Report Review Committee of the National Academy of Sciences. Distribution of the report is approved by the President of the Academy upon satisfactory completion of the review process.

LIBRARY OF CONGRESS CATALOGING IN PUBLICATION DATA

National Research Council. Transportation Research Board.
Freeway entry, flow, and control.

(Transportation research record; 533)

1. Traffic engineering–Congresses. 2. Express highways–Congresses. I. Title. II. Series.
TE7.H5 no. 533 [HE332] 380.5'08s [387.3'1] 75-20287
ISBN 0-309-02384-X

CONTENTS

FOREWORD v

OPERATIONAL ANALYSES OF
FREEWAY MOVING-MERGE SYSTEMS
Samuel C. Tignor 1
Discussion
Joseph A. Wattleworth 16
K. G. Courage 17
Herbert L. Crane 19
Author's Closure 20

PRETIMED SIGNAL-CONTROL SYSTEM FOR
AN UNDERWATER VEHICULAR TUNNEL
S. A. Smith and E. C. Carter 22

DETECTING FREEWAY INCIDENTS UNDER
LOW-VOLUME CONDITIONS
Conrad L. Dudek, Graeme D. Weaver, Gene P. Ritch, and
Carroll J. Messer 34
Discussion
Joseph A. Wattleworth 43
Joseph M. McDermott 44
Harold J. Payne 46

DEVELOPMENT AND EVALUATION OF INCIDENT-DETECTION
ALGORITHMS FOR ELECTRONIC-DETECTOR SYSTEMS
ON FREEWAYS
Masami Sakasita and Adolf D. May 48

EVALUATION OF A PROTOTYPE
WARNING SYSTEM FOR URBAN FREEWAYS
Conrad L. Dudek, R. Dale Huchingson, and Gene Ritch 64

CORQ—A MODEL FOR PREDICTING FLOWS
AND QUEUES IN A ROAD CORRIDOR
Sam Yagar 77

SIMPLIFIED APPROACH TO MODELING FREEWAY
OPERATIONS AND CONTROL
Brian L. Allen and C. Jen Liew 88

MATHEMATICAL EVALUATION OF TRAFFIC CONTROL
ALTERNATIVES FOR RESTRICTED FACILITIES
E. A. Gonzalez, R. C. Loutzenheiser, and E. C. Carter 97

DESIGN TECHNIQUE FOR
PRIORITY-ACCESS RAMP METERING
Saul D. Miller and Harold J. Payne 110

DEVELOPING FREEWAY
PRIORITY-ENTRY-CONTROL STRATEGIES
Khosrow Ovaici, Roger F. Teal, James K. Ray, and Adolf D. May 122

SPONSORSHIP OF THIS RECORD 138

FOREWORD

This RECORD will bring traffic engineers and operators up-to-date on a broad cross section of recent research on improving freeway operations. The papers are on experiments with advanced control systems, research on incident-detection and warning systems, and analytical studies of freeway design and operation.

Extensive research has been sponsored by the Federal Highway Administration during the past few years on moving-merge systems. Although they are not needed for most on-ramps, they can improve operations where ramp geometrics are substandard. Tignor presents findings on driver reactions to 2 types of moving-merge systems. Drivers using the systems improved their merging position without disrupting freeway traffic. Discussions by Wattleworth, Courage, and Crane present additional data on a later application of a moving-merge system and treat operational aspects of the systems.

The earliest controls to improve continuous traffic flow were applied on the trans-Hudson tunnels in New York. For the Baltimore Harbor Tunnel, Smith and Carter report that even a simple control system can significantly benefit traffic flow. Their research on a pretimed signal-control system should help to make this type of system a more practical engineering tool.

Incident detection is a major purpose of surveillance and control systems, and much work has been done in recent years on the logic for incident detection. Dudek et al. consider the incident-detection problem under low-volume conditions. Two computer algorithms were developed. One used a time-scan process; the other, which is considered to be superior, used an event-scan principle. Relations among detector spacings, traffic-flow levels, and incident-detection performance were studied by computer simulations. Discussions by Payne, McDermott, and Wattleworth underscore the difficulty of applying this approach. Imperfections in detector performance would have a major impact on system accuracy, and they propose alternate approaches. This remains an area in which continuing research is needed.

Sakasita and May address the incident-detection problem for all flow levels by means of computer analysis. After reviewing existing detection algorithms and developing 2 new algorithms, they simulate freeway traffic performance and evaluate the algorithms. They compare the modified California detection logic, considered to be the best algorithm in operational use, to the 2 new algorithms. The modified California algorithm performed better at medium and high levels of traffic flow (1,000 and 1,600 vehicles per hour per lane respectively), but the new algorithms were better at low flow (400 vehicles per hour per lane). The study is particularly interesting in its exploration of the influence of detection spacing and flow levels on incident recognition.

When an incident occurs, motorists approaching the scene should be warned that abnormal traffic conditions exist ahead. Dudek, Huchingson, and Ritch report on field experiments on this problem when stoppages occur downstream of cresting vertical curves. Accidents were measured before and after a prototype warning system was installed. A questionnaire survey also was used. The system was found to be cost-effective because accidents were reduced and motorists judged the system to be useful.

Both Yagar and Allen and Liew treat entire freeway systems. Yagar describes a procedure for predicting traffic conditions in a road corridor in which time-varying traffic controls can be simulated and their effects can be explored. This procedure can help traffic engineers to evaluate various control strategies before specialized hardware is installed. The model is also useful as a training aid for students designing control schemes that may involve ramp closure, ramp metering, restriping, and changed traffic signal splits. Allen and Liew propose a simplified approach to this problem by assuming the entire corridor can be represented by only 2 routes: the freeway and 1 alternate route interconnected by equally spaced entrance and exit ramps.

The model is considered particularly useful for initial evaluations of possible control schemes.

Gonzalez, Loutzenheiser, and Carter report on a method for evaluating control alternatives on restricted facilities. They apply the method to the Baltimore Harbor Tunnel. One of 6 possible traffic concentrations measured at each of 3 points in the tunnel defined 18 possible states of tunnel traffic. The state transition probabilities then were determined under each of 5 control alternatives. By this analysis, which used the Howard policy-iteration technique, the most effective control strategy can be defined.

Miller and Payne present a ramp-metering design technique for the situation in which all ramps have a priority-vehicle access lane. The technique minimizes total passenger travel time in a limited corridor. The authors consider the inducement for motorists in nonpriority vehicles to shift to priority vehicles and thereby save time. For the example selected, the San Diego Freeway, this inducement appeared to be weak.

Ovaici, Teal, Ray, and May propose a more general simulation model. This model first predicts freeway traffic performance as a function of freeway design and allowable ramp inflows, then it selects a control strategy that will maximize either the number of persons served or the number of passenger-miles traveled subject to constraints. (For example, the demand for each freeway section will not exceed the capacity of that section.) The model has been applied in San Francisco, Los Angeles, and New York. Typical results indicate that in Los Angeles a 10 percent occupancy shift would reduce vehicle miles of travel by 4.2 percent and enable a 41 percent decrease in the vehicle hours expended by the present demand. Ovaici, Teal, Ray, and May provide a comprehensive treatment of this important subject and illustrate well the usefulness of computer simulation in the development of freeway-control systems.

—Robert S. Foote

OPERATIONAL ANALYSES OF FREEWAY MOVING-MERGE SYSTEMS

Samuel C. Tignor, Federal Highway Administration, U.S. Department of Transportation

Experimental moving-merge control systems were tested in Woburn, Massachusetts. This paper presents findings relative to how ramp drivers used the system and what they thought of the moving-merge concept. A green-band and a pacer system were evaluated. Analyses relative to system effectiveness and use included driver responses obtained from questionnaires, the extent to which ramp drivers used ramp-side displays, and the effect moving-merge systems had on traffic operations. Questionnaire responses indicated that drivers approved of the moving-merge concept and 70 percent found the systems understandable. This statistic was independent of driver age and type of system. Of the drivers who used both systems, 70 percent stated that the green-band system helped most in merging and was easier to understand and use. Analyses were developed to evaluate driver use of the ramp-side, displayed information. These analyses showed that the probability of drivers' using a lighted display downstream of the ramp was more significant for the green-band drivers who had been conditioned to having a lighted display upstream of the ramp. According to the average number of displays viewed per driver, the green-band system was used more consistently than was the pacer system. The mean relative velocity between green bands and ramp vehicles was significantly lower for drivers using the bands. Analyses were used to evaluate traffic operations within the freeway right lane and acceleration lane. Drivers using the systems improved their merge position without disrupting freeway traffic.

•RAMP-CONTROL systems, many types of which have been in use since 1960, control the flow of vehicles onto a freeway to maintain freeway operations at an acceptable level of service. Ramp-control systems can be used on individual on-ramps or on sequences of on-ramps. The most common types of ramp control are total-ramp-closure, pretimed, gap-acceptance, and traffic-responsive systems. Another type of ramp control is the moving-merge system, which uses gap-acceptance control and information display. The displayed information helps the ramp driver identify gaps and merge easily into the freeway flow even when the view of the right lane is restricted. The moving-merge concept was first considered and tested in 1968 by simulated freeway tests on an abandoned airport (1, p. 232). These prototype tests showed that (a) ramp drivers could follow the displayed gap information presented on the ramp, (b) drivers were placed in a successful merge position 70 percent of the time, (c) the moving-merge concept was feasible, and (d) further development was warranted.

After these tests, the Federal Highway Administration contracted to develop functional requirements, control logic, and design specifications for 2 types of moving-merge systems. On the basis of the design specifications, each system was fabricated, assembled, and operated at a single on-ramp in Woburn, Massachusetts (2, pp. 4-11; 3). The Woburn site was selected for the first public tests because it has a long, 700-ft (213-m) acceleration lane. This long acceleration lane offered a safety advantage by providing additional space for ramp vehicle maneuvering.

The design of both merge-control systems was based on gap-acceptance control. The green-band system represented right-lane gaps as moving green bands on a ramp-side display. The pacer system used a green pacer light to lead ramp drivers to the merge area. The green-band and pacer systems are shown in Figures 1 and 2. The green-band system used open-loop control in which trajectories of the green bands were independent of the behavior of the ramp vehicles. The pacer system, however, used closed-loop control, and the individual pacer-light trajectories were based on the movement of both the freeway gap and the ramp vehicle.

The basic purposes of the Woburn evaluation were to determine whether the development and operation of moving-merge systems would be technically feasible, how ramp drivers would react to the moving-merge concept, and whether ramp drivers would use the system.

SYSTEM DESCRIPTIONS

Experimental field tests were performed on each system in 1970 at Routes 38 and 128 in Woburn. For both systems a Raytheon 703 minicomputer with 12,000 words of core storage was used for surveillance, decision-making, and system control. Peripheral equipment included a high-speed, paper-tape unit for reading operational programs and a magnetic tape unit for recording real-time data. For surveillance and evaluation, 7 sets of inductive loop sensors 200 ft (60.8 m) apart were used to monitor vehicle movement in the freeway's right lane; 10 sets of sensors 64 ft (19.2 m) apart were used on the ramp. Five 50-ft (15.1-m) presence sensors were installed in the acceleration lane to monitor the presence of stopped vehicles.

Green-Band System

The green-band system operated in (a) moving, (b) stopped-gap-acceptance (SG), and (c) stopped-metered (SM) states. In the Woburn tests, the state in which the green-band system operated was determined by the average 3-min speed in the freeway's right lane. When the average 3-min speed was greater than 35 mph (56 km/h), the system operated in the moving state.

In the moving state, the computer determined the location of acceptable gaps in the right lane as each vehicle crossed an inductive-loop detector in the freeway. Representations of these acceptable gaps were then displayed as moving green bands on the ramp-side display unit. When the ramp driver stayed adjacent to a green band as it moved at a constant speed along the ramp-side display, he or she would arrive in the merge area within an acceptable gap. The moving green bands lengthened, shortened, or disappeared depending on how the right-lane gaps varied.

When the average 3-min speed fell below 35 mph (56 km/h), the green-band system operated in either the SG or SM state. The SG state was similar to conventional, pretimed, ramp-control systems except that in it the traffic signal released a waiting ramp vehicle with a green indication and a 32-ft (9.7-m) accelerating green band on the

Figure 1. Green-band system.

Figure 2. Pacer system.

ramp-side display when an acceptable gap was available. The short green band would lead the ramp vehicle to the acceptable gap in the merge area. If no acceptable gap were found within a predetermined time period, the waiting vehicle would be released but no accelerating green band would be used. In the SM state, all vehicles were metered individually without an accelerating green band.

Pacer System

The major difference between the pacer system and the green-band system was that in the pacer system the speed and location of all ramp vehicles were continuously monitored as they moved along the ramp. As in the green-band system, the computer maintained a list of right-lane vehicle arrivals in the merge area. When a vehicle entered the ramp, the computer calculated its expected time of arrival in the merge area and searched the freeway list for a gap large enough for the ramp vehicle. When the computer matched a ramp vehicle with an acceptable gap, an individual pacer light was displayed to the driver. The pacer light, positioned about 1 car length in front of the driver, guided the ramp driver to the merge area so that both the ramp driver and the freeway gap arrived at the same time. More than 1 ramp vehicle could be accommodated simultaneously. The movement of each pacer light was accelerated or decelerated by the computer according to the relative relationship between the ramp vehicle and its respective merge gap. For the Woburn tests, ramp vehicles that lost an acceptable gap or those that did not have an acceptable gap continued along the ramp without a pacer light.

RAMP DRIVER RESPONSES

The green-band and pacer merge-control systems were operated from 7:00 a.m. to 7:00 p.m. for approximately 8 weeks each. The pacer system operated during May and June 1970, and the green-band system operated during September and October 1970. Before the start of each operational period, local newspapers described how the systems functioned and how drivers should use the information. In addition, brochures describing how the systems were to be used were distributed at the on-ramp for several days before each phase.

The success or failure of traffic aids such as the green-band and pacer merge-control systems depends on drivers. To appraise individual driver reactions to the merge-control systems, questionnaires were distributed to the ramp drivers after each system had been in use for 8 weeks. The questionnaires were distributed at 3 locations near the on-ramp entrance in the morning, at midday, and in the afternoon. The morning period represented the peak freeway and ramp flows. The off-peak flows occurred during the midday and afternoon periods. One thousand five hundred and twenty green-band system questionnaires and 1,582 pacer system questionnaires were distributed. Nineteen percent of the green-band questionnaires and 25 percent of the pacer questionnaires were returned.

System Clarity

Table 1 was prepared to determine whether drivers responded differently, on the basis of type of system or driver age, to the question, "Was the entire merging control system clear and understandable?" With respect to driver age, the chi-square analysis shows that age is independent of type of merge-control system and system clarity. The degrees of freedom (df) for independence of age, A, systems, S, and clarity, C, is 10. The df was obtained by subtracting $(A - 1) + (S - 1) + (C - 1)$ from $(A \times S \times C) - 1$.

In addition to examining the independence of age, type of merge-control system, and system clarity, we examined other interactions. The chi-square test can be used to calculate the individual interactions. But, as Kullback (5, pp. 12-14) has shown, information theory also can be easily employed for tests of contingency tables; the information statistic is distributed asymptotically as is chi-square, and it has additive and convex properties that make its use convenient. Therefore, information statistics for each of the interactions are given in Table 2. In addition, the information statistics for all of the main effects as well as a comparative chi-square statistic for each main effect and interaction are included in Table 2.

As shown by the Table 2 interactions, neither the information statistic nor the chi-square test rejected independence among age, system clarity, or type of merge-control system. Consequently, there was no degradation in system clarity because of driver age or type of merge-control system. Also the only main-effect component of interest is system clarity; about 70 percent of the drivers found the system clear and understandable.

System Location

Merge control is intended to be used at substandard ramps where drivers experience difficulty in merging and where reconstruction of the merge area is not feasible or is too costly.

The data of Table 3 were used to test the independence of the observed response for the 2 merge concepts. Respondents were asked in the questionnaires whether they thought the systems were needed at most ramps, needed at poorly designed ramps only, or not needed. As indicated in Table 4, the chi-square test did not reject the hypothesis of independence between the 2 data sets. This suggests that the drivers' opinions on where merge control was needed were not significantly affected by the type of merge system considered. Furthermore, about half of the drivers stated that the systems were needed at poorly designed ramps only.

System Difficulty

In order to evaluate the drivers' views of their experiences in using the merging systems, 2 questions on the degree of difficulty in driving beside the display lights were asked. The first concerned their first use of the systems and the second concerned their use after gaining experience. Degrees of difficulty were difficult, slightly difficult, and easy. The responses for the green-band and pacer merging systems are given in Table 4.

A test was made to determine whether acquaintance with the system was independent of type of merge system and degree of difficulty. As shown by Table 4, the hypothesis of independence was rejected, implying that the observed percentages at each level were not uniform for each system from one time to another. Based on this finding, the statistical significance of both the main effects and interactions was computed by using information theory. Results are given in Table 5.

Before reviewing the components of Table 6, one should understand that no restriction was placed on the number of green-band and pacer questionnaires analyzed. The number analyzed was solely a function of the number of usable questionnaires returned for each system. Consequently, no real importance can be associated with the significant main effects for the type of system, S, or for acquaintance with system use, T; similarly, no real importance can be associated with the significant interaction of $T \times S$. The remaining significant components shown in Table 6 now can be discussed in greater detail.

For the main effect on level of difficulty, D, the information statistic clearly shows that the distribution of responses at each level, combined over systems, was highly significant; approximately 65 percent of the drivers answered that it was easy to drive beside the moving-display lights.

The interaction of acquaintance with the system and level of difficulty, $T \times D$, was significant. This implies that the drivers detected a difference in how difficult it was to drive beside the moving-display lights for the first time as compared to how difficult it was after they had gained experience. It can also be shown that the same conclusion

Table 1. Effect of driver age on system clarity.

Age (years)	Response[a] Yes Number	Yes Percent	No Number	No Percent	Sample
Green-Band System					
Under 22	13	61.9	8	38.1	21
22 to 30	47	62.7	28	37.3	75
31 to 55	105	71.4	42	28.6	147
Over 55	11	68.8	5	31.2	16
Total	176	68.0	83	32.0	259
Pacer System					
Under 22	18	78.3	5	21.7	23
22 to 30	49	64.5	27	35.5	76
31 to 55	111	74.0	39	26.0	150
Over 55	11	73.3	4	26.7	15
Total	189	71.6	75	28.4	264

Note: $\chi^2 = 5.89$, $\chi^2_{0.05,\,10} = 18.3$.

[a]The question was, "Was the entire merging control system clear and understandable?"

Table 2. Main effect and interaction components of Table 1.

Component	Information	χ^2	Degrees of Freedom	$\chi^2_{0.05,\,df}$
Main effects				
A	345.748	348.182	3	7.81
C	84.215	81.930	1	3.84
S	0.045	0.047	1	3.84
Interactions				
A × C	3.940	4.010	3	7.81
A × S	0.112	0.123	3	7.81
C × S	0.821	0.820	1	3.84
A × C × S	0.976	0.977	3	7.81
All effects and interactions	435.857	436.089	15	25.00

Table 3. Where merge control is needed.

System	Suggested Location: Most Ramps Number	Most Ramps Percent	Poorly Designed Ramps Only Number	Poorly Designed Ramps Only Percent	Not Needed Number	Not Needed Percent	Sample Size
Green band	62	24.8	115	46.0	73	29.2	250
Pacer	91	26.2	169	48.7	87	25.1	347
Total	153	25.6	284	47.5	160	26.9	597

Note: $\chi^2 = 1.26$, $\chi^2_{0.05,\,2} = 5.99$.

Table 4. Degree of difficulty of driving beside moving-display lights.

Acquaintance With System	Level of Difficulty: Difficult Number	Difficult Percent	Slightly Difficult Number	Slightly Difficult Percent	Easy Number	Easy Percent	Sample Size
Green band							
First	25	16.0	59	37.8	72	46.2	156
Experienced	8	7.1	34	30.1	71	62.8	113
Total	33	12.3	93	34.6	143	53.1	269
Pacer							
First	23	6.8	121	36.0	192	57.2	336
Experienced	16	4.9	65	20.1	243	75.0	324
Total	39	5.9	186	28.2	435	65.9	660

Note: $\chi^2 = 58.61$, $\chi^2_{0.05,\,7} = 14.1$.

is valid when the green-band and pacer responses are considered individually. This is supported by the fact that the information statistics for the individual green-band and pacer analyses were 9.10 and 24.16 respectively, which, when compared to $X^2_{0.05}$ with 2 df, rejects the hypothesis of independence. Drivers had less difficulty in using the systems after they had gained experience.

The information statistic also showed that a significant interaction for D × S was present. Closer review of Table 4 indicates that the overall ranking of level of difficulty in driving beside the moving display was higher for the pacer system than for the green-band system. This means that drivers found it easier to drive beside the display in the pacer system. For example, nearly 15 percent more pacer responses were marked easy. This is probably because the pacer system updated the movement of the pacer light according to how the ramp driver was moving along the ramp.

Preferred System

On the green-band questionnaire, those drivers who had used both systems were asked the questions given in Table 6 to determine which system was preferred. Table 6 determines whether the number of individual preferences for both of the 2 systems were homogeneous across each query. As given in Table 6, the chi-square test failed to reject homogeneity among the 3 system queries. This can also be shown by the normal approximation of the binomial test that preference was for the green-band system and that the proportion, p, of drivers having a preference for the green-band system can be expressed as $Pr(0.67 < p < 0.77) = 0.95$. Approximately 70 percent responded that the green-band system helped most in merging and was easier to understand and to use.

USE OF RAMP-SIDE DISPLAY

Two approaches were developed to determine how drivers used the ramp-side displays. One approach considered the composite use of the display by all ramp drivers regardless of whether they had a choice in using the display information. The second approach was more selective in that it considered only those drivers who could choose whether to use the display information.

Composite Measures of Display Use

The measures described here were composites in the sense that they included all drivers who used the ramp regardless of the chance they had to use the moving-merge information. Thus the composite measures can be used to compare the effectiveness of the 2 systems with respect to each other as well as to no system.

Conditional Lighted Display Probabilities

The driver display, located directly beside the vehicle, was either lighted or unlighted when a driver crossed a given set of ramp detectors. A lighted display indicated that the driver was under control and was expected to arrive in the merge area at the same time that an acceptable gap in traffic would be available on the freeway. An unlighted display indicated that the driver was not expected to arrive in the merge area at the same time as an acceptable gap in traffic. The conditional probability of a driver's having a lighted display at sensor location j (S_j) if there were one at sensor location i (S_i) is expressed as

$$p(j \mid i) = \frac{p(i,j)}{p(i)}, \quad (j > i)$$

Table 5. Main effect and interaction components for Table 4.

Component	Information	Degrees of Freedom	$\chi^2_{0.05, df}$
Main effects			
Acquaintance with system, T	3.26	1	3.84
Level of difficulty, D	453.17	2	5.99
System, S	169.80	1	3.84
Interactions			
T × D	33.09	2	5.99
T × S	3.86	1	3.84
D × S	16.83	2	5.99
T × D × S	0.18	2	5.99
All effects and interactions	680.19	11	19.70

Table 6. Driver preferences.

Question	Green Band		Pacer		Sample Size
	Number	Percent	Number	Percent	
Which system was easier to understand?	78	70.3	33	29.7	111
Which system was easier to use?	91	72.2	35	27.8	126
Which system helped most in merging?	73	73.7	26	26.3	99
Total	242	72.0	94	28.0	336

Note: $\chi^2 = 0.32$. $\chi^2_{0.05, 2} = 5.99$.

Table 7. Model notation.

System	Notation	Probability
Green band		
On	x_1	p_{11}
Off	x_2	p_{12}
Total	N_1	
Pacer		
On	y_1	p_{21}
Off	y_2	p_{22}
Total	N_2	
Both		
On	$x_1 + y_1$	p_1
Off	$x_2 + y_2$	p_2
Total	$N_1 + N_2$	

Table 8. Green-band and pacer system comparisons of conditional probability of a lighted display at S_j given a lighted display at S_i.

System	S_i	$S_{j=1}$			$S_{j=2}$			$S_{j=3}$			$S_{j=4}$			$S_{j=5}$		
		Volume	p(j|i)	$2\hat{I}$[a]	Volume	p(j|i)	$2\hat{I}$[a]	Volume	p(j|i)	$2\hat{I}$[a]	Volume	p(j|i)	$2\hat{I}$[a]	Volume	p(j|i)	$2\hat{I}$[a]
Green band	1	71	1.00	0	58	0.82	6.35[b]	57	0.80	2.71	53	0.75	1.78	53	0.75	4.84[b]
Pacer	1	292	1.00	0	268	0.92	6.35[b]	206	0.71	2.71	194	0.66	1.78	177	0.61	4.84[b]
Green band	2				85	1.00	0	69	0.81	5.09[b]	62	0.73	1.06	63	0.74	5.16[b]
Pacer	2				353	1.00	0	243	0.69	5.09[b]	237	0.67	1.06	215	0.61	5.16[b]
Green band	3							89	1.00	0	71	0.80	0.72	68	0.76	2.42
Pacer	3							301	1.00	0	227	0.75	0.72	204	0.68	2.42
Green band	4										96	1.00	0	77	0.80	2.65
Pacer	4										309	1.00	0	222	0.72	2.65
Green band	5													104	1.00	0
Pacer	5													287	1.00	0

System	S_i	$S_{j=6}$			$S_{j=7}$			$S_{j=8}$			$S_{j=9}$		
		Volume	p(j|i)	$2\hat{I}$[a]	Volume	p(j|i)	$2\hat{I}$[a]	Volume	p(j|i)	$2\hat{I}$[a]	Volume	p(j|i)	$2\hat{I}$[a]
Green band	1	50	0.70	3.76	50	0.70	4.80[b]	51	0.72	7.34[b]	50	0.70	6.21[b]
Pacer	1	169	0.58	3.76	164	0.56	4.80[b]	158	0.54	7.34[b]	158	0.54	6.21[b]
Green band	2	59	0.69	4.42[b]	60	0.71	5.95[b]	60	0.71	7.11[b]	58	0.68	6.02[b]
Pacer	2	201	0.57	4.42[b]	198	0.56	5.95[b]	193	0.55	7.11[b]	189	0.54	6.02[b]
Green band	3	63	0.71	2.40	63	0.71	2.94	65	0.73	6.17[b]	62	0.70	5.20[b]
Pacer	3	186	0.62	2.40	183	0.61	2.94	176	0.58	6.17[b]	169	0.56	5.20[b]
Green band	4	72	0.75	4.38	67	0.70	1.86	66	0.69	3.19	63	0.66	1.52
Pacer	4	196	0.63	4.38	192	0.62	1.86	181	0.59	3.19	181	0.59	1.52
Green band	5	88	0.85	4.94	83	0.80	4.95[b]	81	0.78	5.73[b]	76	0.73	3.40
Pacer	5	212	0.74	4.94	196	0.68	4.95[b]	187	0.65	5.73[b]	181	0.63	3.40
Green band	6	107	1.00	0	87	0.81	1.27	84	0.79	2.14	82	0.77	2.66
Pacer	6	270	1.00	0	205	0.76	1.27	192	0.71	2.14	184	0.68	2.66
Green band	7				108	1.00	0	94	0.87	5.94[b]	91	0.84	7.25[b]
Pacer	7				272	1.00	0	206	0.76	5.94[b]	193	0.71	7.25[b]
Green band	8							110	1.00	0	103	0.94	16.46[b]
Pacer	8							270	1.00	0	204	0.76	16.46[b]
Green band	9										114	1.00	0
Pacer	9										248	1.00	0

[a]This statistic obeys the approximate χ^2 distribution with 1 df.

[b]Significantly different at the 0.05 level.

This analysis uses the concept of conditional probability to determine whether use of the driver display differed for the green-band and pacer systems. For the ideal situation, if a ramp driver were always under control and had a display at detector i, then the probability $p(j|i)$ that he or she would be under control at detector j always would be 1.00.

In the resulting model for each i,j combination $j > i$ [with underlying $p(j|i)$], 2 independent random samples, N_1 and N_2, corresponding to 2 system conditions, were obtained. Each of these samples thus can be regarded as a set of independent observations from a binomial distribution. The samples N_1 and N_2 were for the green-band and pacer systems respectively; a complete notational description is given in Table 7.

Each observation in the set of N_1 observations is assumed to have a probability p_{11} (that the ramp driver had an "on" display at location i and an "on" display at location j). Thus, in this model, $p_{11} \equiv p_1(j|i)$ and $p_{12} \equiv 1 - p_{11}$. Similarly for set N observations, $p_{21} \equiv p_2(j|i) \equiv 1 - p_{22}$. We wish now to determine whether $p_{11} = p_{12}$. The null hypothesis to be tested is that samples $x = (x_1, x_2)$ and $y = (y_1, y_2)$ are from the same overall population, $p = (p_1, p_2)$. Kullback (5, p. 128) analyzed this type of problem in terms of the information statistic 2I for the more general case involving c data sets. Kullback showed that 2I can be approximated by

$$2I = \frac{1}{N_1N_2}\sum_{k=1}^{c}\frac{(N_2x_k - N_1y_k)^2}{x_k + y_k}$$

with c - 1 df.

Table 8 gives the data and set of analyses for each location corresponding to a merge volume of 1,750 vehicles per hour (vph). For appropriate interpretation, it must be realized that the 2I statistics, calculated for each value of i, are not independent. That is, if the value of 2I at $j = 2$ for $i = 1$ is low, one would expect that a similar result would prevail at $j = 3$ for $i = 1$ because of the proximity of these locations. Even if the 2I statistics were independent, one would anticipate under the null hypothesis that 5 percent of a large quantity would be higher than the critical 0.95 value. Twenty 2I statistics for 1,750-vph merge volume and twenty-four 2I statistics for 1,000-vph merge volume exceeded the critical 0.95 value of chi-square. Because numbers as large as these are quite unlikely, the green-band and pacer conditional probabilities are not homogeneous. Studied interpretation of the results can show that the probability of a driver's having a lighted display at location j was higher for the green-band system than for the pacer system if a lighted display was at i.

In examining the difference in these results, one should remember that the pacer system was a closed-loop system designed to lead and keep ramp drivers within an acceptable gap. If the pacer concept had been truly effective, the $p(j|i)$ would have been greater for the pacer system than for the green-band system. However, the findings suggest otherwise. The results of the conditional probability studies suggest that drivers tended to stay within an acceptable gap more often with the green-band system than with the pacer system. For the pacer system, only 1 pacer light was displayed per driver at any time; consequently, drivers did not have as extensive a visual impression of the size of the gap as they had with the green-band system. The limited information provided from the pacer light might be the key explanation for the difference in these results.

Average Number of Lighted Displays per Driver

A ramp driver who approached the beginning of the driver display with the merge-control system in operation may or may not have had a lighted display. If a lighted display was at each detector along the length of the ramp, he or she used 1 uninterrupted green band or pacer to drive in relation to only 1 acceptable freeway gap. If, on the other hand, the gap closed and was no longer acceptable, the lighted display was stopped by the system somewhere along the ramp. When this occurred the ramp driver might have obtained another green band or pacer for a different acceptable gap at some point further along on the ramp. Thus, a ramp user may drive beside a succession of k uninterrupted moving displays ($k = 0, 1, 2, \ldots$) while moving down the ramp to the merge area. No display light signified that the computer was unable to find an acceptable gap for the driver.

When the freeway merge volume was zero, the average number of continuous green bands or pacers that would be viewed by the ramp driver would be 1 because, at this volume, the driver would have an uninterrupted lighted display throughout the length of the ramp. At the other extreme, as the freeway merge volume approached capacity (2,000 vph), the average number of green bands or pacers that would be displayed to a ramp vehicle would approach zero. At freeway merge volumes between zero and capacity, the average number of green bands or pacers used by a ramp driver varied depending on (a) the probability of a green band or pacer and (b) how drivers used the 2 systems.

To determine whether there was any difference in how drivers used the ramp-side information for the 2 systems, an evaluation was made based on the comparison of 2 independent data sets. This method hypothesizes that the average number of lighted displays per driver, $\bar{k}$, is the same for both the green-band and pacer systems. The comparison required the use of least squares regression models. The following models were used:

$$y = 1 + a_1 m + a_2 m^2$$

$$y^* = 1 + b_1 m + b_2 m^2$$

where

y = average number of continuous displays per driver for the green-band system,
y^* = average number of continuous displays per driver for the pacer system,
$a = (a_1, a_2)$, a least squares estimate of the model parameter,
$b = (b_1, b_2)$, a least squares estimate of the model parameter, and
m = merge volume in 100 vehicles per hour (when $m = 0$, $y = y^* = 1$).

Individual parameter values are given in Table 9.

The comparison also required that a least squares estimate [$c = (c_1, c_2)$] be determined for the combined green-band and pacer sets of data. After all parameters were determined, an F-ratio was computed. The result of this analysis is given in Table 10, and, based on the F-statistic obtained, the above hypothesis was rejected at the 0.05 level of significance. Thus it can be inferred that, because $\bar{k}$ was less for the green-band drivers (1.3) than it was for the pacer drivers (1.5), the green-band display was used more consistently.

Selective Measures of Driver Display Usage

Because the selective measures to be described apply to either type of merge-control system and because future installations probably will be green-band systems, the analyses in this section have been restricted to the green-band system.

Some ramp drivers had no opportunity to use the moving green band as they moved down the ramp. This happened for 2 reasons. First, the driver arrived at the ramp entrance when there was no acceptable gap in traffic on the freeway and there never would be an acceptable gap no matter how much he or she accelerated. Second, the driver sometimes arrived at the ramp entrance when there was such a large gap in traffic that, no matter how much he or she accelerated or decelerated, he or she would

Table 9. Model parameters.

System	Data Set	Parameter 1	Parameter 2
Green band	On	0.0660	-0.00290
	Off	0.0606	-0.00256
Pacer	On	0.0690	-0.00221
	Off	0.0651	-0.00185

Table 10. Differences in display use measured by average number of lighted displays per driver for drivers having at least 1 display.

System	Data Set (i)	Data Points (N_i)	Sum of Squares[a] Regression (RSS_i)	Sum of Squares[a] Total (TSS_i)	Sum of Squares[a] Differences (R_D)	Sum of Squares[a] Error (ESS)	F-Ratio[b] [$F_{0.05,(2,df)}$]
Green band	1	7	12.310	12.314			
Pacer	2	11	22.810	22.893			
Both	1 and 2	18	35.120	35.207	0.100	0.087	8.046[c] (3.74)

[a] $R_D = RSS_1 + RSS_2 - RSS_{1\ and\ 2}$ and $ESS = TSS_{1\ and\ 2} - RSS_1 - RSS_2$.

[b] $df = N_{1\ and\ 2} - 4$ and F-ratio $= \frac{R_D/2}{ESS/df}$.

[c] Results differ significantly at the 0.05 level.

Figure 3. Green-band and vehicle time-space plot.

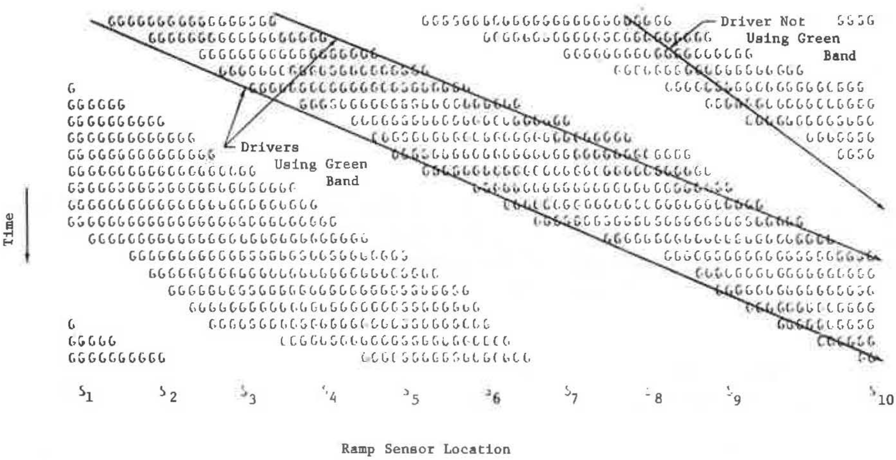


Table 11. Homogeneity of system-on and system-off, green-band, time-space data.

System Status	Driver Had No Chance To Get Out of Green Band: Number	Percent	Driver Had No Chance To Get Into Green Band: Number	Percent	Driver Had Chance to Use Green Band: Number	Percent
On	119	25.1	190	40.1	165	34.8
Off	89	29.3	119	39.1	96	31.6

Note: $\chi^2 = 1.82$, $\chi^2_{0.05,2} = 5.99$.

Table 12. Frequency comparison of drivers who may have used a green band as a function of system-on and system-off status.

System Status	Drivers Who Used Green Band: Number	Percent	Drivers Who Did Not Use Green Band: Number	Percent
On	111	67.3	54	32.7
Off	39	40.6	57	59.4

Note: $\chi^2 = 17.63$, $\chi^2_{0.05,2} = 3.84$.

always be beside the same gap. These types of cases were not studied so that a detailed study could be made of those vehicles that had a chance to use the displayed information.

The data for these analyses were obtained from green-band and vehicle time-space plots. Examples of typical time-space plots are shown in Figure 3. Each 8-ft (2.4-m) section of an individual green band is designated as G. The trajectories of the vehicles moving along the ramp are represented by a solid line. Vehicle trajectories on the time-space plots were examined to identify those for which the driver had a choice in the green-band display. A driver was considered to have had a choice in using a green band if he or she could have negotiated beside a green band before reaching the downstream end of the display without having to accelerate more than 4 ft/sec^2 (1.2 m/s^2) and if the following 2 conditions were not rejected:

1. The driver was able to see the moving green band in front of or beside his or her vehicle, and
2. The driver was not able to choose a green band because his or her movement was impeded by another ramp vehicle.

Probability Comparisons

The data for this analysis were obtained while the system software was operational but while the display was either on or off. With the data structured in this way, it was possible to compare how drivers used the ramp when the system was on to how they used it when it was off. In this analysis, merge volumes were sought that would permit a large number of green bands to be generated and observed; the merge volumes were between 900 and 1,500 vph.

Prior to the usage analysis, it was considered appropriate to determine whether the system-on and system-off data sets were similar with respect to the categories of drivers who (a) had no chance to get out of a breen band, (b) had no chance to get beside a green band, and (c) had a chance to use a green band. The chi-square test was used to evaluate the homogeneity of the system-on and system-off samples. As shown in Table 11, the chi-square test did not reject homogeneity for composition of the number of system-on and system-off observations for each of the 3 categories.

After finding that the 3 categories were similar, 2 individual evaluations were performed on the data set to determine the frequency with which drivers used green bands and how they used the green bands at the beginning and end of the display. First, a chi-square test was used to determine whether the degree of driver's use of the green-band system was homogeneous with the system's status. It is shown in Table 12 that the chi-square test rejected homogeneity between on and off. When the system was off, 41 percent of the drivers who had a chance to use the system performed in such a way that they appeared to use the undisplayed information, or, in other words, the system coincided with the driver's behavior. However, when the system was on, 67 percent of the drivers who had a chance to use the system did indeed use the displayed information. This clearly indicates that drivers were using the system. Second, an evaluation considered the display status when ramp drivers were at the beginning and at the end of the driver display for all those drivers who had a chance to use a green band. Transition matrices were developed for the system-on and system-off cases. These matrices are given in Table 13. Kullback, Kupperman, and Ku (6) have shown that information theory can be used to determine whether several realizations of matrices of transition probabilities are homogeneous. Their method was used for this evaluation and the results, given in Table 14, indicate that, for those drivers who had a chance to use a green band, the probability of their having a green band at the end of the driver display was improved significantly when the display was energized. The finding also suggests that some drivers were using the green-band system. The significance of this improvement is represented by the significant chi-square value of 6.2521 for the component effect of conditional homogeneity $(k|j)$. The j homogeneity component of information merely indicated that the composition of states, disregarding the system-on and system-off conditions, was homogeneous and extraneous to our findings.

Table 13. Transition matrices for drivers who could have used the green-band display.

System Status	Beginning of Display (j)	End of Display (k): Not Beside Green Band, Number	Not Beside Green Band, Percent	Beside Green Band, Number	Beside Green Band, Percent
On	Not beside a green band	25	36	44	64
	Beside a green band	31	32	65	68
Off	Not beside a green band	21	43	28	57
	Beside a green band	25	53	22	47

Table 14. Homogeneity of transition matrices of Table 13.

Component	Information	Degrees of Freedom	$\chi^2_{0.05,\,df}$
j homogeneity	2.0814	1	3.84
Conditional homogeneity k|j	6.2521	2	5.99
j, k homogeneity	8.3335	3	7.81

Table 15. Green-band and relative velocity for drivers who used and drivers who did not use a green band.

Driver Use of Green Band	Sample Size	Cumulative Distribution (fps): -15	-10	-5	0	5	10	15	20	25
Drivers used green-band system	84	0.01	0.08	0.21	0.53	0.83	0.90	0.95	0.99	1.0
Drivers did not use green-band system	42	0.02	0.07	0.14	0.31	0.57	0.74	0.83	0.95	1.0

Note: 1 fps = 0.3 m/s.

Table 16. Gap-acceptance distributions.

Case	System Status	Display Status	Variance of Sample	Mean (sec)	Critical Gap[a] (sec)	Variance of Mean	Standard Normal Deviation
1	Off	—	10.431	3.584	3.5	0.250	—
2	On	Used	17.215	4.678	4.8	0.178	1.673[b]
3	On	Not used	10.946	3.795	3.6	0.233	0.304[c]

[a]Critical gap as determined by the Raff method.
[b]Mean is not significantly different from the mean of case 1 at the 0.10 level.
[c]Mean is significantly different from the mean of case 1 at the 0.10 level.

Table 17. System-off versus system-on leading headways.

System Status	Driver Usage of Band	Sample Size	Cumulative Distribution (sec): 1	2	3	4	5	6	7	8	>8
Off	—	125	0.088	0.376	0.544	0.688	0.752	0.800	0.864	0.904	1.00
On	Used	105	0.048	0.181	0.352	0.543	0.657	0.791	0.819	0.867	1.00

Table 18. Ratios of relative velocity to highway vehicle velocity for green band, system on, and green band, system off.

System Status	Sample Size	Cumulative Distribution (fps): -0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4	0.5
Off	125	0.01	0.02	0.08	0.17	0.36	0.57	0.76	0.94	0.00	1.00
On, used	105	0.00	0.01	0.04	0.13	0.31	0.57	0.77	0.95	0.98	1.00

Note: 1 fps = 0.3 m/s.

Driver Tracking of a Moving Display

The task of following a moving display is analogous to following another vehicle. For vehicle-following cases, Montroll and Potts (7, p. 43) found that relative velocity between the 2 vehicles is a major influence on tracking. It thus can be hypothesized that the relative velocity between the green bands and the ramp vehicles would be less for drivers who used the moving display than for drivers who did not use the display. As used here, the relative velocity is the difference between the green-band and ramp-vehicle velocities. The analysis considers 2 subsets of drivers who had a chance to use green bands. One subset included the drivers who used green bands, and the other included those drivers who did not use a green band. Table 15 gives the cumulative distribution of the observed relative velocity for those drivers using and those not using a green band. The critical deviation would be 0.25, and the actual maximum absolute deviation between 2 distributions was 0.26 and occurred at a relative velocity of 5 fps (1.5 m/s). So, the distributions were significantly different at the 0.05 level. For this observed difference, the Kolmogorov-Smirnov 2-sample test would reject the hypothesis that the 2 samples came from the same distribution.

THE EFFECTS OF MOVING-MERGE SYSTEMS ON TRAFFIC OPERATIONS

The ability of ramp vehicles to merge into freeway traffic is affected by both geometric design elements and traffic operations at the site. The geometric design elements normally are fixed for a given site and cannot be altered except through reconstruction. Modest site alterations or improvements, including moving-merge systems, can affect traffic operations. The use of moving-merge systems to aid ramp drivers in merging into the freeway's right lane falls into this category. Thus evaluation of the effectiveness of moving-merge systems must involve the various traffic operational variables or measures that are featured.

Gap-Acceptance Characteristics

Drivers in a moving-merge system can time their arrival at the merge area to correspond with an acceptable gap. When there is no moving-merge system or when freeway traffic is hidden because of visual obstructions along the ramp, a driver would not have premerging information. Intuitively, then, one might expect that the mean gap accepted for merging would be different for the 2 cases. In fact, because the green-band system does not display small gaps (gaps of less than 2 sec), it would be expected that when drivers used the system the mean gap would be larger than when the system was not used because drivers would accept a greater percentage of small gaps. This section considers 1 case in which no green-band system was operated and 2 cases in which the system was operated and was either used or not used by the ramp drivers.

For this analysis, a method developed by Karber is used because it provides the mean, standard deviation, and variance of the mean of a set of increasing observed proportions where observed proportion is the percentage of drivers accepting a given-sized gap (8, p. 10.3; 9, pp. 201-202).

Statistical tests were made to determine whether the mean of the accepted gaps for the cases in which the green-band system was operated was significantly different from the case in which the green-band system was not operated. By use of the normal approximation relative to 2 independent data sets, one can determine whether the mean of set 1 is significantly different from the mean of set 2. The results of these tests are given in Table 16 along with critical gaps found by the Raff method. The use of the normal approximation did not permit rejection of the hypothesis that the mean gap for the system-off case is equal to the mean gap for the system-on case when drivers do not use the display. However, the normal approximation did permit rejection of the hypothesis, at the 0.10 level, that the mean gap for the system-on case when drivers use the display is equal to the mean gap for the system-off case. Thus these findings confirm the statement that the green-band system had a significant effect, at the 0.10 level, on the gap size selected by drivers.

Leading Headways After Merge

For this analysis, leading headway is defined as the headway between the ramp vehicle and the upstream highway vehicle. If drivers used moving-merge systems, the leading headway would be greater when the system was used than when no merge system was available because, with the use of the merge-control system, a leading headway allowance is applied to each highway vehicle in the freeway's right lane. As a result, the displayed green band is shorter than the actual freeway gap. The Kolmogorov-Smirnov 2-sample test was used to test the hypothesis that the distribution of leading headways is the same for the system-on and system-off cases. Based on the results given in Table 17, the evidence supports the statement that leading headways were significantly larger for that system-on case when drivers used the system than for cases when the green-band system was not used. The critical deviation would be 0.179, and the actual maximum absolute deviation between 2 distributions was 0.195. So, the distributions were significantly different at the 0.05 level.

Ratio of Relative Velocity to Highway Vehicle Velocity

Another consideration relevant to the merge process is the relative velocity between the highway and ramp vehicles involved in the merge. Drew (10, p. 201) defined a model that related time gap, T, between 2 highway vehicles upstream of the merging area to time, T′, that the ramp driver perceives as the gap length. His expression is

$$T/T' = \frac{V - V_r \cos \alpha}{V}$$

where

T/T' = ratio of actual gap to the perceived gap,
V = velocity of the upstream highway vehicle,
V_r = velocity of the ramp vehicle, and
α = angle at which the ramp and freeway converge.

This model can be used as a measure for evaluating the effect the merge-control system had on the relative speeds between merging vehicles. For the Woburn site α was 5.70 deg.

The Kolmogorov-Smirnov 2-sample test given in Table 18 failed to reject the hypothesis at the 0.05 level that the 2 distributions came from the same population. The critical deviation would be 0.179, and the actual maximum absolute deviation between 2 distributions was 0.05. So the distributions were not significantly different at the 0.05 level. Consequently, evidence suggests that at the 900 to 1,500-vph merge volume the relative behavior between the ramp and freeway vehicles was not changed when the green-band system was used at the site.

CONCLUSIONS

This report presented the results of an evaluation of freeway merge-control systems tested in Woburn, Massachusetts, in 1970. The Woburn tests were designed to determine the technical feasibility of freeway merge-control systems and whether drivers approved of the concept and used the ramp-side driver displays.

It was found that a moving-merge system was technically feasible, and the results of the analyses showed that ramp drivers approved of the concept, recognized that the systems would be used only at poorly designed ramps, and preferred the green-band system over the pacer system. It was found that drivers tended to stay within an acceptable gap more often with the green-band system than with the pacer system, and the green-band display was more consistently used by ramp drivers. Results were presented that showed that the variance of the relative velocity between the green-band edges and vehicles was significantly greater for those drivers who did not use a green band than for those drivers who used a green band. This finding also provided evidence that some drivers did use the moving green bands.

The analyses showed that the mean gap for the system-off case was significantly smaller than it was for the system-on case and that the leading headways were significantly larger when the green-band system was operated and drivers used the system than when the system was not used. Thus it was found that those ramp drivers who elected to use the green-band system and who drove beside a green band improved their positions within the freeway gap; however, they did not alter either their behavior within the acceleration lane or the behavior of the freeway traffic from what it would have been if no system had been available. That is, there was merge improvement and no traffic disruption.

These analyses are suitable for future traffic-operation evaluations and, in particular, future merge-control systems such as the green-band system being tested in Tampa, Florida.

ACKNOWLEDGMENTS

This research was part of a staff research project at Office of Research and Development, Federal Highway Administration. Appreciation is expressed to Donald E. Cleveland of the University of Michigan, who, as dissertation advisor, offered numerous suggestions and recommendations to me during the preparation of this paper. Special thanks also are due to Sidney Weiner of the Federal Highway Administration, for his suggestions of statistical analyses.

REFERENCES

1. D. R. Drew et al. Gap Acceptance and Traffic Interaction in the Freeway Merging Process—Phase II Final Report. Texas Transportation Institute, Texas A&M Univ., 1968.
2. Pacer System Analysis and Design. In Merging Control System, Raytheon Co., Sudbury, Mass., Vol. 1, 1970.
3. Green Band Design. In Merging Control System, Raytheon Co., Sudbury, Mass., Vol. 8, 1970.
4. S. C. Tignor. Operational Analyses and Improvements for Freeway Moving-Merge Systems. Univ. of Michigan, PhD dissertation, 1974.
5. S. Kullback. Information Theory and Statistics. John Wiley and Sons, Inc., New York, 1959.
6. S. Kullback, M. Kupperman, and H. H. Ku. Tests for Contingency Tables and Markov Chains. Technometrics, Vol. 4, No. 1962, pp. 602-604.
7. E. W. Montroll and R. B. Potts. Car Following and Acceleration Noise. HRB Spec. Rept. 79, 1964, pp. 37-48.
8. M. G. Natrella. Experimental Statistics Handbook 91. National Bureau of Standards, U.S. Department of Commerce, 1963.
9. J. D. Church and B. E. Cobb. On the Equivalence of Spearman-Karber and Maximum Likelihood Estimates of the Mean. Jour. American Statistical Association, Vol. 68, March 1973.

10. D. R. Drew. Traffic Flow Theory and Control. McGraw-Hill Book Co., New York, 1968.

DISCUSSION

Joseph A. Wattleworth, Department of Civil Engineering, University of Florida

Tignor has conducted a significant analysis of a freeway moving-merge system, and his results and conclusions have far-reaching effects. The nature of the study should be kept in mind when the results are considered. The 2 types of moving-merge control systems—the pacer and green-band systems—were installed at the Woburn site primarily to evaluate the hardware. The site was selected to minimize the geometric and traffic problems to allow the evaluation of the electronic operation of the systems.

The traffic-operation evaluation of the systems was a secondary aspect of the installation, and Tignor has done well in this evaluation. He compared the 2 systems and presented the following major conclusions:

1. There was no significant difference of driver opinion on the clarity of the pacer and green-band systems. About 70 percent of the drivers found the systems clear and understandable.
2. Over 70 percent of the drivers found it easy to drive beside the moving displays after they had some experience in using the systems. Experience helped the drivers understand the systems and use them properly.
3. About 70 percent of the drivers who used both systems preferred the green-band system and indicated that it was the more helpful and was easier to understand.
4. Drivers tended to stay within an acceptable gap more often with the green-band system than with the pacer system.

Thus the results of the studies that were conducted indicated that the green-band system was operationally superior to the pacer system. This is especially significant considering the fact that the cost of the green-band system is much lower than the pacer system because it requires fewer detectors on the ramp. In addition, the conceptual appeal of the open-loop, green-band system with the display representing an acceptable gap is greater to many traffic engineers than is that of the pacer system. Thus there is no need to trade off cost against effectiveness in comparing the systems. The green-band system not only is more effective but also is cheaper.

There was an apparent inconsistency in the results of the questionnaire studies that were conducted to compare the systems. A questionnaire study was conducted to determine driver evaluation of the ease of use of the 2 systems.

When the green-band system was installed, drivers who used both systems were asked to indicate which system was easier to use. The percentage of the pacer questionnaire respondents that indicated that the pacer system was easy to use was higher than the corresponding percentage of the green-band questionnaire respondents. However, in the questionnaire study conducted on the drivers who had used both systems, a greater percentage indicated that the green-band system was easier to use.

A comparison was made of the placement of ramp vehicles relative to the green with and without the display of green bands. It was found that, when no green bands were displayed, 40 percent of the ramp drivers who were in a position to possibly use the green band were positioned in such a way that they would have been beside a green band if it had been displayed. When the green bands were displayed, 67 percent of the drivers who were in a position to possibly use the green band were found to be beside a green band. Thus the green-band system can be considered to have improved the merge position and, consequently, the merge operation of 26 percent of the ramp vehicles that potentially were able to use the green band.

At first consideration, it may seem that providing a benefit to less than 26 percent of the ramp drivers is not significant. But several operational measures of effectiveness (such as delay and merge accidents) are quite sensitive to the probability of merging without a stop. Thus an increase on the order of 26 percent in merges without stopping would have a great effect not only on merge delay but also on probability of a rear-end collision in the merge area.

It was found that the average accepted gap was higher for vehicles that used the green-band display than for (a) vehicles that did not use the green-band display or (b) all vehicles when the green bands were not displayed. This would suggest that the gap portion of the program should be changed to yield lower required gaps for green-band generation. It raises a potential trade-off that must be made in evaluating the ability of the green-band system—the gap size versus the probability of merge success. By providing green bands for only relatively large gaps, the probability of a successful merge will be higher, as will the resultant safety of the merge operation, but the merge capacity will be lower. This will increase the tendency of the system to revert to a stopped mode and would increase queuing and delay on the ramp.

The results presented in the paper were based on studies conducted at a location at which merge and geometric problems were minimal. It will be interesting to compare them to the results of the evaluation of the green-band system in Tampa, Florida, where the operational and geometric problems are more severe.

In summary, the author has presented a very thorough analysis of 2 alternative approaches to a new type of freeway-ramp, moving-merge control. The results are of great practical value and provide a basis for evaluations that are being conducted in Tampa.

K. G. Courage, University of Florida

Tignor has done an excellent job of describing the evaluation of the first experiment with moving-merge control in Woburn, Massachusetts. My comments will deal primarily with a subsequent experiment with the same system that took place on the Ashley Street entrance ramp to I-75 in Tampa, Florida. The comments fall into 2 general categories: those dealing with system changes and refinements necessary to operate the green-band system in Tampa and those dealing with a preliminary evaluation of this system.

SYSTEM MODIFICATIONS

The Woburn system was developed with considerable flexibility to change operating parameters in anticipation of future deployment at other entrance ramps. It was, of course, necessary to modify detector-location parameters and various system-calibration parameters to adapt the system to the geometrics of the Tampa ramp.

Some important changes also were made to the display hardware. The green-band display itself was improved significantly through the use of a continuous row of fluorescent units as opposed to a series of discrete incandescent lamps. This change was made possible primarily because of the more favorable temperature conditions that are found in Tampa. It is not known how well fluorescent units would perform in a colder climate. Some changes in the driver information signing also were instituted in Tampa. The advanced displays that used 2 fixed-message signs with flashing signals in Woburn were consolidated into a single-lamp-matrix, changeable-message sign installed over the ramp as a gateway to the system. The sign says DRIVE BESIDE GREEN BAND whenever the system is in the moving-merge mode. In all other operational modes PREPARE TO STOP is displayed. The sign is blank when the system is inoperative. As another departure from the Woburn system, the MERGE WITH CAUTION blank-out sign in the merge area was replaced with a blank-out yield sign whose dimensions and color conform to the manual on uniform traffic devices. It also was necessary to make a number of operational software changes to adapt the system to the Ashley Street ramp. The Woburn ramp is more or less linear, but the Ashley Street ramp has a fairly sharp curve [230-ft (70.1-m) radius] on which the maximum safe speed is approximately 30 mph (48 km/h). Therefore, the propagation characteristics of the band to travel at 30 mph (48 km/h) for the first half of the ramp and accelerate to the discharge velocity of 45 mph (72 km/h) by the end of the display had to be changed. This modification also did away with the advisory speed sign indicating the constant speed at which the green band was traveling, and the elimination of the constant-speed green bands may have destroyed some of the potential benefits of merge control.

An additional change was made to the system-control logic to improve mode selection. The Woburn system was programmed to restart whenever congestion was detected in the merge area. This was found to cause problems in Tampa because of the higher level of merge-area activity. The response to merge-area congestion therefore was changed to simply mask the green bands but continue the remainder of the computational processes. Restarting is now carried out in response to an ambiguous condition arising from detector data. The conditions for changing from the moving mode to the SG mode also were altered. The Woburn system changed modes whenever the speed in the freeway right lane crossed a predetermined threshold. It was found in Tampa that because of ramp geometrics, heavy merge volumes did not cause an appreciable drop in freeway speed and that low speeds generally were observed during adverse weather only. Merge volume as measured just downstream of the entrance ramp proved to be a much more reliable parameter for this purpose. Separate volume thresholds were provided to control transition in each direction. This introduced a hysteresis effect that made the system more stable.

PRELIMINARY EVALUATION

The system has been fully operational in Tampa for approximately 2 months. Data collection and analysis are now in progress. The current status of the analysis does not permit definite conclusions to be drawn about its effectiveness. Observation, however, indicates that the system is able to generate green bands that flow into gaps in the freeway traffic and that motorists try to use the system.

On the negative side, the propagation of the band down the ramp appears to be less smooth than is desirable because of differences in speed projections obtained from the various sensors throughout the system. More refined computational algorithms would be required to improve this situation. Speeds on the ramp appear to have increased somewhat probably as a result of the green bands. However, deceleration from the end of the band display to the merge area also has increased. Approximately 20 percent of the vehicles have shown decelerations greater than 8 ft/sec^2 (2.4 m/s^2). This may suggest a lack of confidence on the part of the driver.

More definite conclusions will be based on 4 types of data to be collected and analyzed. Operational data are being collected daily from the detectors. These data will provide such information as speeds, flows, and accelerations. Limited, human-factor studies also are being carried out with test subjects in an instrumented vehicle. A public questionnaire will be distributed on the ramp to determine motorists' reactions to the system. Finally, accident reports for a before-and-after period will be analyzed to assess what, if any, improvements are attributable to the system. To date, no accidents have been recorded in which the system can be considered at fault.

Based on the limited analyses performed, I offer the following preliminary conclusions that are based largely on opinion and reflect only my views:

1. Following the green band in the moving mode of operation is not a particularly difficult task; however, some improvement in the propagation characteristics of the band (for example, elimination jerkiness) is desirable.
2. Use of the green band in the SG mode appears questionable. The distance from the stop line to the merge area is fairly short, and the acceleration characteristics of

individual drivers and vehicles vary too much to benefit substantially from an acceleration profile based on an average vehicle. The occasional absence of a green band because of the absence of a suitable gap on the freeway causes some confusion.

3. The present green-band operation is strictly open loop, that is, the propagation characteristics are not based on any information obtained from the ramp so the green bands will frequently overtake vehicles that are traveling very slowly or are stopped on the ramp. This creates a definite potential for rear-end collison on the ramp. I feel, therefore, that some feedback of detector information is required to promote safer ramp operations.

4. The complexity of the system has caused some operating and maintenance problems. System operation is extremely sensitive to malfunctioning of the input detectors, and erratic performance from any of the approximately 25 detectors can cause a complete shutdown.

Although some of these comments appear to be negative, I feel that the overall concept of the system is sound. With some mechanical improvements and sufficient driver education, the moving-merge approach to ramp control could be developed into a practical and workable traffic-control system. The question of cost effectiveness, however, is likely to limit its application to only a relatively small number of critical ramps in congested urban areas.

Herbert L. Crane, Michigan Department of State Highways and Transportation

Tignor has been comprehensive in his analysis and has done a fine job in accomplishing his stated objectives. It is understandable that, for safety, a test site with conservative geometrics was selected. I feel that the system-off data on the type of ramp for which these systems were intended would be further from the optimum that existed at the Woburn location. The reader is left with the impression that a greater contrast would exist between the system-off and system-on data at a geometrically more complex location. Even greater benefits would then be shown for that type of installation.

A questionnaire polled drivers' impressions of the need for the system (all ramps, poorly designed ramps, or not at all). Although it is impossible to determine what factors the driver considered in forming an impression, the driver was sure to notice that the quantity and the sophistication of the equipment were much greater than he or she normally encountered in similar situations. The driver also could have easily concluded that the system was very expensive. I have a strong impression that the drivers, by considering only an apparent cost rather than a consideration of cost versus benefits, would be inclined to be negative in their answers.

I wonder whether the level of difficulty in driving beside the display refers to the hardware or to the strategy because both of these parameters were changed at the same time. This may imply the need for a third candidate system, such as one with green-band hardware and some of the pacer system's updating logic.

The use of questionnaires for base data dictates that the driver respond to the questions at some time after his or her performance in the system. The longer the wait before the driver fills out the questionnaire, the more prone he or she is to have lapses of memory or mistaken impressions. It would be interesting to measure the driver's reactions to the systems by visual evidence of compliance. I recognize, however, that to supplement Tignor's work with such a study would be costly, time consuming, and difficult to analyze.

The provision for a leading headway allowance, resulting in a shorter green band, is obviously necessary for safety. It is possible that this provision might have introduced the driver onto the freeway at a headway other than that which he or she usually considered comfortable. This would suggest that some drivers may have responded to the questionnaire in a negative way even though they benefited by the system.

The indication that relative behavior between the ramp and freeway traffic in the merge area was not changed by the green-band system suggests that the conservative site was an optimal design for the traffic conditions encountered. I would expect stronger differences between system-off and system-on statuses to occur on more critical ramps.

Tignor has stated that merge control is intended to be used at substandard ramps where drivers experience difficulty in merging and where reconstruction of the merge area is infeasible. He has addressed himself to the problem adequately for the intended scope of his work. I feel that his statement deserves support and emphasis because some of the applications are less apparent than others. The most apparent application is its use on ramps with critical sight restrictions, short merge lanes, and other similar features. Less apparent are those in which the aforementioned features are adequate but contain other geometric features that make the ramp drivers' evaluation of freeway gaps difficult to evaluate. The best example of this would be a circular ramp of minimum radius. In this case the driver must remain intent on the task of tracking around the curve and, at the same time, try to find and evaluate a gap in traffic on a roadway to which the angle of approach is constantly changing. All the while, the driver must keep his or her vision fixed on the ramp, which is curving to the right, and the gap on the freeway, which has a motion toward the left periphery of his or her vision because of the rotational component of his or her own motion. The introduction of a moving-merge guide display close at hand tends to bring the visual references into a more narrow and constant field of vision to which the driver can more readily relate.

AUTHOR'S CLOSURE

I would like to thank Wattleworth, Courage, and Crane for their reviews of the moving-merge system tested in Woburn, Massachusetts. The points they made will be very helpful in providing further improvements in the moving-merge concept. It was particularly interesting to learn from Courage's remarks that the preliminary results from the Tampa questionnaire were similar to the results obtained in the Woburn tests. Several specific points were raised by the reviewers that require additional discussion.

Wattleworth stated the percentage of respondents to the pacer questionnaire (the first system tested indicating that the pacer system was easy to use was higher than the corresponding percentage for the green-band system. These results were taken from the question concerning the degree of difficulty in using the moving-display lights. For 2 reasons the individual system questionnaire results should not be jointly compared. First, not all drivers used both systems, and second, the basis of how individual drivers determined level of difficulty was not uniform for each system. For example, a driver may have answered the question for the pacer system while subjectively comparing it to no system. For the green-band questionnaire, a driver could have subjectively compared the green-band system to both the pacer system and no system. The last question of the green-band questionnaire was included to permit a system-to-system comparison. This question is believed to provide the most accurate information on ease of use because only those drivers who used both systems were considered.

Courage stated that the propagation of the band is sometimes interrupted or disturbed because of differences in speed projections obtained from the various sensors. He suggests that a refined green-band algorithm is needed to stabilize the propagation of green bands along the driver display. This suggestion appears to have considerable merit, and it should be considered for future system installations.

Courage also stated that the preliminary results indicated that the ramp speeds have increased but that there also had been some increase in deceleration from the end of the display to the merge area. He further stated that 20 percent of the drivers might have been electing to slow down or not merge when they had a green band. It will be interesting to see after several months of operation what effect additional driver use and system operation will have on this preliminary finding.

The SG mode was designed to permit drivers to relate to the location of the freeway gap and thereby more readily merge into it. Courage's comments suggest that the acceleration profile being used in Tampa requires drivers to accelerate either faster than they desire or faster than their vehicle will permit. A slower acceleration profile may be necessary. If a suitable acceleration profile cannot be determined that a reasonable number of vehicles can use, it may be best, as Courage suggested, not to use the SG mode. However, it is my view that the SG mode can be very helpful and should be maintained as a driver aid.

The green-band operation uses an open-loop control concept. Courage expressed some concern about green bands overtaking slowly moving ramp vehicles and suggested that some feedback mechanism be included. At present system feedback and overriding provisions are included in the green-band software. For example, whenever vehicles are stopped within the acceleration lane for longer than 5 sec, the display is blanked out until the acceleration lane is clear. This safety override appears sound when one considers that drivers do not lock into moving green bands but monitor other ramp vehicles.

One of the provisions included in the Tampa green-band evaluation plan provides for a cost-effectiveness analysis. This analysis will identify those applications in which the use of the green-band system will be the economically preferred plan. Experience in Tampa indicated that an operating moving-merge system may cost between $125,000 and $150,000 with an annual maintenance and operation cost of $5,000 to $10,000.

PRETIMED SIGNAL-CONTROL SYSTEM FOR AN UNDERWATER VEHICULAR TUNNEL

S. A. Smith,* JHK and Associates; and
E. C. Carter, University of Maryland

Previous research has indicated that limiting traffic demand in an underwater vehicular tunnel can produce substantial increases in flow rates and speeds. A pretimed control system consisting of a standard traffic signal was used to meter traffic entering the Baltimore Harbor Tunnel. Four different cycle lengths and splits were tested, and the results were compared to those obtained from the uncontrolled situation. The increase in flow rates and speeds of 3 of the control strategies over those in normal operation showed that even a simple control system is capable of significantly benefiting traffic flow through a tunnel. An evaluation methodology was applied to the control alternatives; both quality of flow and flow rates were considered. Primary emphasis was placed on a methodology requiring a minimum of time and effort in data collection. The operation of the tunnel was observed in its 2 basic states—congested and uncongested. The quality of flow in a tunnel is poorest on the downgrade, better on the level section, and best on the upgrade. The results of this research should help to make pretimed, tunnel-bottleneck control a more practical engineering tool.

•BOTTLENECKS on urban streets and freeways greatly affect the flow of traffic. The rate of flow on a highway can be no higher than the flow at its most critical bottleneck. Many bottlenecks, such as those due to construction, are temporary, and the quality of flow can be restored to an acceptable level in a relatively short period of time. Other restrictions, such as underwater tunnels, are permanent fixtures, and operational improvements must improve the level of service.

It is apparent from a review of the literature that many underwater tunnels in urban areas have insufficient capacity, high accident rates, and frequent vehicle stoppages. The research discussed in this paper was performed to show that these undesirable conditions can be greatly improved with a minimal amount of capital investment and technical skill.

It has been found in past research that, because of the geometric configuration of underwater tunnels, a bottleneck frequently exists near the point where the grade changes from level to an upgrade (1). Because vehicles tend to decelerate when they first start to climb the upgrade, the vehicular concentration increases upstream from this point, and further decreases in speed and flow take place. This effect continues upstream until the entire line is moving slowly. This stop-and-go type of movement not only reduces the flow rate through the bottleneck but also increases output of pollutants, fuel consumption, the probability of a collision, and driver frustration. Furthermore, when the concentration has risen to the point where flow deteriorates, it is very difficult to regain a state of efficient flow and lower concentrations as long as the demand remains high.

It was the purpose of this research to determine whether a pretimed signal-control

*Mr. Smith was with the University of Maryland when this research was performed.

system is capable of significantly increasing both the quality of flow and flow rates through a tunnel faced with this type of grade-induced bottleneck. The northbound lanes of the Baltimore Harbor Tunnel were selected as the study site. As will be shown, pretimed signal control, when properly used, can achieve this goal, and a simple and inexpensive tool will be readily available to the traffic engineer. Previous research has shown that the tunnel proper is the critical link in the Baltimore Harbor Tunnel complex (2).

Although no comprehensive summary of research on restricted facilities could be found, Everall (3) has provided an excellent summary of both past research and the state of the art in traffic surveillance and control, particularly for urban freeways.

METHOD OF STUDY

Site Description

The Baltimore Harbor Tunnel, a section of the Harbor Tunnel Thruway, is a major link in the Northeast Corridor that carries not only intercity traffic but also a significant portion of the commuting traffic in the Baltimore area. Like the thruway, it consists of 2 lanes in each direction. The southbound lanes carry the greater proportion of traffic in the morning peak, and the northbound lanes carry the greater proportion in the evening peak. Lane changing is prohibited throughout the tunnel, and trucks are restricted to the right lane. MAINTAIN 45 signs are posted on the approach and within the tunnel.

The plan and profile views of the tunnel are shown in Figure 1. The major northbound problem is the 4 percent grade change at the beginning of the upgrade. Therefore, the tunnel offers an excellent opportunity to study the effect of grades and the effect of grade-induced bottlenecks on traffic flow.

Data Acquisition System

Vehicle Detector System

The Baltimore Harbor Tunnel now is equipped with an extensive data collection system monitoring the northbound lanes. Seven stations in each lane, consisting of 2 photocell detectors slightly over 13.5 ft (4.1 m) apart, are capable of sensing many desired flow characteristics. Use was made of frequency-division multiplexing and frequency-shift keying concepts in the design of the data-transmission system. The location of the stations can be seen on the tunnel profile in Figure 1. More details on this system are available elsewhere (2).

Quality of Flow Data

Because an improvement is likely to gain public approval if the driver notices it, better ride quality must be a main objective in a bottleneck control system. A 1965 Chevrolet equipped with a Greenshields traffic analyzer was driven in the traffic stream to evaluate the effects of traffic control through the eyes of the driver. Runs were made in the early morning hours to obtain a base free-flow condition in average off-peak situations and during peak hours for each alternative to evaluate the relative effects of each control strategy. Smith and Carter (4) describe the operation of the analyzer in greater detail.

Design of Experiment

Data collection using the system of photocell detectors began in February 1973. Included in the first set of data were peak-period flows (3:30 p.m. to 6:30 p.m.) for a Tuesday, Thursday, and Friday. An additional day of Tuesday data was collected in March. This produced 12 hours of traffic characteristics for the uncontrolled situation.

The next phase was the installation of the traffic metering system. One traffic signal with 12-in. (0.3-m) signal lenses was installed in each of the 2 northbound lanes in November 1973. They are located on the downgrade approaching the tunnel, approximately 1,200 ft (366 m) upstream of the entrance. Signs were placed at various locations upstream of the signal to warn the motorist of the possibility of being stopped. The purpose of the signal was to periodically introduce gaps into the traffic system in a manner similar to that used in studies on New York tunnels (5, 6). It was felt that this would disperse some of the long vehicle platoons that commonly form in the peak periods and cause speeds to decay. These gaps were created by programming a short red time into the signal cycle. To observe the effects on traffic of very restrictive and slightly restrictive control strategies, several different cycle lengths and splits were used (Table 1). The results were compared to the no-control data.

The cycle lengths were initially selected by comparing the probable capacity of the bottleneck with the capacity at the signal for various cycle lengths. The capacity for each cycle length was determined by calculating the theoretical maximum number of time headways that can be contained within 1 green indication at the signal. This was expanded to an hourly volume and compared to the expected hourly volume for the foot of the upgrade. The calculation of the alternative cycle lengths was not overly critical, however, because cycles and splits could be changed easily if, after the first day of control, it became obvious that some other control strategy was needed.

Metering generally was begun between 4:00 and 4:30 p.m. The signal was activated only when traffic was congested from the bottleneck to a distance approximately 1,000 to 1,500 ft (305 to 457 m) upstream of the signal. This meant that speeds were approximately 20 mph (32.2 km/h) as vehicles passed the signal and that the potential for rear-end collisions due to the signal was reduced. After the first 5 days of data collection, it was decided that stopping traffic at the signal for 90 sec to allow the tunnel to be cleared of congested traffic would permit each alternative to seek its own traffic state without outside influence. This was then done by a tunnel officer at the beginning of each data collection period.

The vehicle with the traffic analyzer was driven through the tunnel at the same time that the metering was done. The course started approximately 1,000 ft (305 m) upstream of the signal and was completed at the exit portal of the tunnel. Flow parameters, including elapsed time, distance, and change in velocity, were printed every 5 sec. It was felt that this would produce changes in velocity in small enough time increments to accurately calculate the mean velocity gradient, the parameter used to measure quality of flow.

Because the comparison of alternatives was the primary objective, the same driver was used for all runs. This eliminated any difference in driving techniques. Runs were made in both the right and left lanes. It was usually possible to make 6 runs each day during the data collection period. The resultant data from the detectors and analyzer were then processed and analyzed.

EVALUATION METHODOLOGY FOR ALTERNATIVE CONTROL STRATEGIES

Before a tunnel administrator selects pretimed signal-control strategies, he or she must determine which strategy yields the best results for a given traffic condition. The purpose of the methodology developed in this study was to provide the administrator with procedures to use that will result in a good decision.

Figure 1. Detector locations, plan and profile views of the Baltimore Harbor Tunnel.

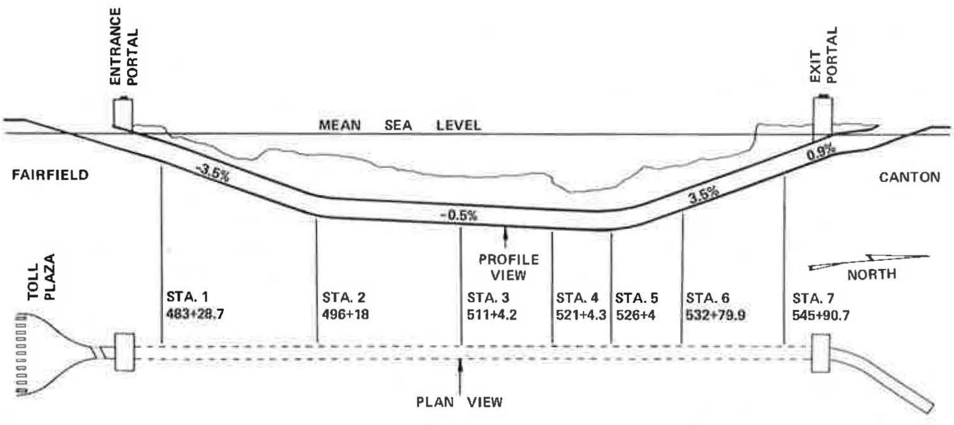
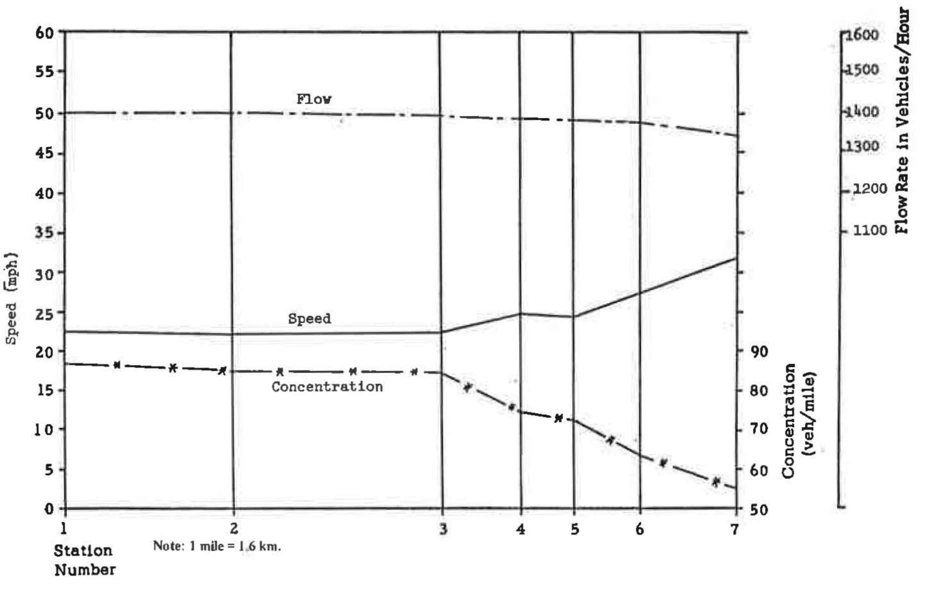


Table 1. Cycle lengths and splits.

Cycle Length (sec)	Green Time (sec)	Red Time (sec)	Amber Time (sec)	Green + Amber Time (percent)
120	109.2	7.2	3.6	94
160	147.2	8.0	4.8	95
180	169.2	7.2	3.6	96
240	225.6	9.6	4.8	96

Figure 2. Traffic-flow profile, Feb. 15, 1973, left lane, uncontrolled.



Location of the Bottleneck

This is the first step in the technical phase of implementing and evaluating a tunnel control system. In many situations in which congestion occurs, the bottleneck is readily identified by visual inspection. A queue of vehicles usually indicates the presence of a bottleneck at the front of the queue. For instance, a traffic signal acts as a bottleneck on its red indication.

Previous research showed that the toll plaza acts as a bottleneck when demand is less than about 2,800 vehicles per hour (vph)(2). This can be seen from the short queues and delays that occur during off-peak hours. When demand begins to exceed 2,800 vph, the location of the bottleneck is less obvious. It was a normal occurrence before the early 1974 gasoline shortage for vehicles to back up from the tunnel to the toll plaza during peak hours so that drivers often had to wait in the toll booth even after they had paid their toll, which indicated that the bottleneck was somewhere downstream of the toll plaza.

Observations of flows inside the tunnel indicate that the bottleneck cannot always be located by examining queue lengths alone because vehicles are constantly moving and concentrations are constantly changing. One might be led to believe that the horizontal curve at the tunnel exit could be the bottleneck.

Through a statistical analysis, Palaniswamy (7) showed that speed is a good parameter to use to identify a bottleneck. Any point in a congested system at which speeds are increasing is likely to be downstream from a bottleneck. As can be seen from the speed profile of the tunnel during congested conditions (Fig. 2), the increase in speed more precisely identifies the beginning of the tunnel upgrade as the critical bottleneck. There is a slight increase in speed after station 3, but there is a major increase after station 5, which indicates that the foot of the upgrade is the critical bottleneck. Figure 2 also includes profiles of flow and concentration. When the bottleneck in a congested tunnel cannot be determined by visual inspection, a speed profile should be helpful.

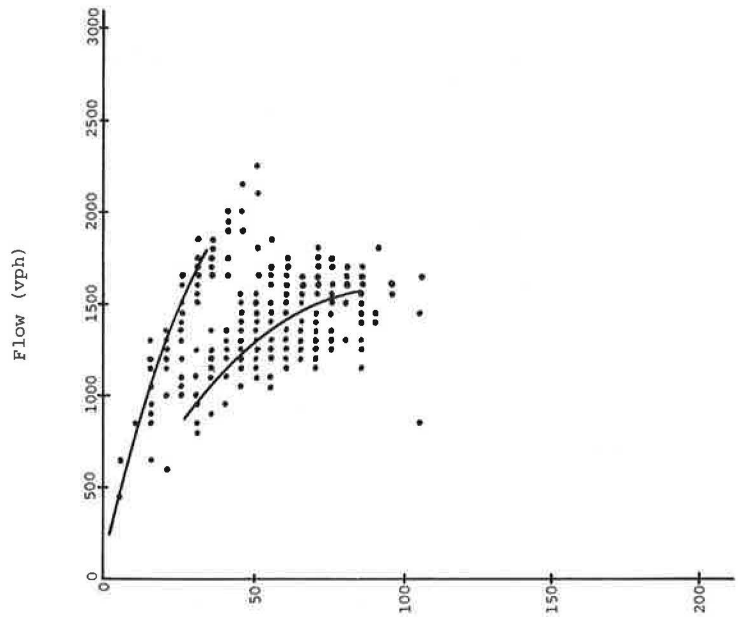
Capacity of the Bottleneck

According to theory, the point of zero slope of the flow-concentration (q-k) curve is the capacity of the roadway. A question arises, however, when there is a discontinuity in the data representing the bottleneck as shown in Figure 3 (4). Although much further analysis must be done to validate this point, it appears from Figure 3 that there may be 2 q-k curves representing traffic flow at the bottleneck. The tendency toward 2 curves is indicated by the 2 lines drawn through the data points. This presents a rather confusing picture when one is trying to determine the capacity of the bottleneck. An average of the data points at the apex of the lower curve would indicate a capacity of approximately 1,500 vph. Some of the data points on the upper curve indicate that higher flow rates are obtainable. To resolve this question a deeper analysis was undertaken.

Because the slope of the chord drawn from the origin to any point on the curve indicates the speed corresponding to the flow rate and concentration for that point, it is evident that speeds on the upper curve are higher than those on the lower curve. A histogram of 30-sec speeds (Fig. 4) suggests that speeds in the middle range [28 to 35 mph (45.1 to 56.4 km/h)] occur much less frequently than do those of either high or low range. This lack of middle-range speeds corresponds to the area between the 2 curves in Figure 3 where few data points exist. It appears, therefore, that traffic in the bottleneck region can take on basically 2 different states, which are described by these 2 speed ranges. When concentrations are not yet at the point where speeds break down, average speeds are normally between 35 and 48 mph (56.4 and 77.2 km/h) (state 1). When concentrations become too high, mean speeds drop abruptly into the 18- to 28-mph (29- to 45-km/h) range (state 2). These states correspond to the upper and lower q-k curves respectively.

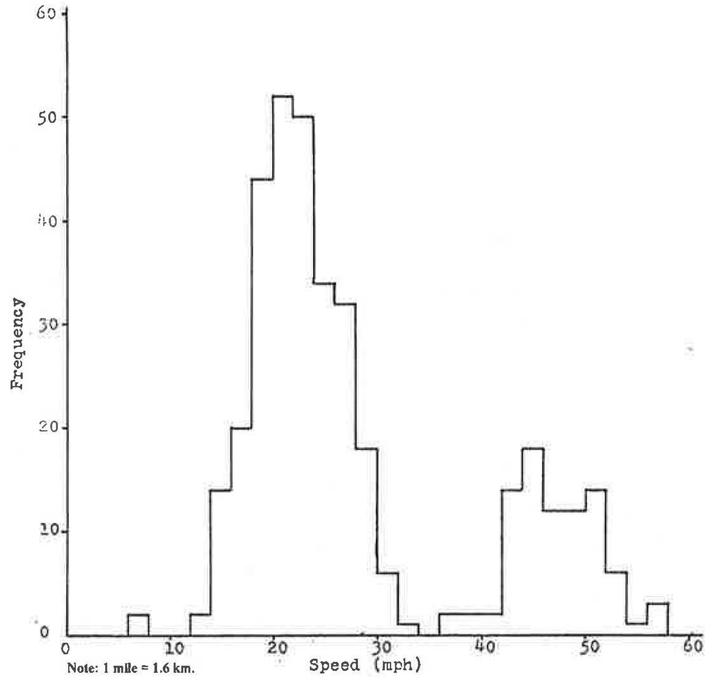
Palmer (8) also suggested the possibility of 2 q-k curves for bottlenecks from a study of a construction bottleneck. As demand increases, the flows may correspond to the

Figure 3. Flow versus concentration, Feb. 13, 1973, station 5.



Note: 1 mile = 1.6 km.

Figure 4. Frequency of 30-sec speeds, Feb. 13, 1973, station 5.



Note: 1 mile = 1.6 km.

high-speed curve (state 1), but, when concentrations reach the critical level, the flow drops into the lower curve described by state 2. Thus it may be possible for flow rates to be at the same level for 3 different combinations of speed and concentration. Because some of the highest flow rates are found in state 1, it would be advisable, for the purpose of tunnel-bottleneck control, to maintain speeds in this range. Maintaining high speeds would yield not only a better quality of flow but also high flow rates that could be maintained longer than they could in state 2.

Palaniswamy (7) has noted that the high kinetic energy of vehicles traversing the upgrade at higher speeds may account for higher flow rates. For instance, a car whose speed has doubled has 4 times the kinetic energy it would have at the lower speed. The additional energy and the motive power of the engine are used to maintain the higher speed throughout the upgrade. Thus the shock waves caused by the deceleration of vehicles may be less severe and occur further downstream. The higher kinetic energy is of special benefit to trucks whose great inertia helps to minimize the loss of speed in climbing. This leads one to believe that flow increases are obtainable in the truck lane as well as in the left lane. The necessity of the driver's being more alert at higher speeds may also contribute to higher flows. Unfortunately, it takes only 1 driver to adversely influence the traffic flowing in a single lane. An abnormal number of these drivers on 1 day may cause flows to be poor even at high speeds. Conversely, flows at low speeds also may be high on some days.

In summary, tunnel-bottleneck flow tends to assume 1 of 2 basic states, which are defined by 2 speed ranges. The flows allowed by the pretimed signal should be set at a level that will cause speeds to remain in state 1. At present, it can be deduced (from experience on the New York and Baltimore tunnels) that for state 1 the capacity of a traffic stream without trucks is about 7 percent higher than the average hourly volume during congestion in a tunnel. A properly operated demand-responsive system should further increase the capacity of the bottleneck. Because a tunnel is unlike an open roadway in geometrics and because of the differing psychological responses of drivers to a tunnel, the Highway Capacity Manual (10) procedure for calculating capacity does not apply to tunnel bottlenecks.

Development of Control Alternatives

When the maximum controlled flow rate of the bottleneck has been determined, control alternatives should be designed. Although the restriction may be at a point other than the foot of the upgrade, previous research in the New York and Baltimore tunnels indicates that the upgrade will be the critical section in most underwater tunnels. If a toll plaza is the problem area, vehicle processing might be improved by the installation of another toll booth. Each tunnel will present its own particular set of problems.

As stated previously, a traffic signal acts as a bottleneck on the red indication. In essence, a traffic signal on an urban street is a controlled bottleneck to alleviate a possibly greater bottleneck and thus increase the capacity of the street system. By introducing a controlled bottleneck into a section of the roadway upstream of the critical bottleneck, flow through that critical bottleneck can be increased. Ideally, the capacity of the control point should be regulated so that it is the same as or just below the capacity of the bottleneck.

Various control methods and strategies have been considered in the past, but, because the traffic signal is a widely accepted control device that is simple and inexpensive to operate, it is highly recommended for tunnel-bottleneck control. Several combinations of cycle lengths and splits should be chosen for initial experimentation. Other strategies may be designed after the first few days of data collection based on observation of traffic flow during the initial strategies.

Analysis of Control Strategies

A major problem with comparing flows achieved by each of the control strategies is that the same drivers do not pass through the tunnel every day, which makes flows dependent on the human factor. Also incidents occur at random, and flow can be interrupted for as little as 2 or 3 min to as long as half an hour. The number and duration of incidents during each alternative testing period may bias the results. The human factor cannot be realistically eliminated, but for evaluation purposes it is assumed not to vary between alternatives. The effect of incidents can be eliminated either by taking data for an extended period of time or by removing all flows affected by incidents. Long data collection periods are not desirable, particularly if several control strategies are to be attempted. Although comparing flow rates by removing incident-related data is slightly more complex, it takes much less data collection time. It also indicates more specifically what the cause of an improvement actually was. If, for instance, daily peak-hour flows, including incidents, are averaged over time, one cannot be sure how much of the improvement is due to control and how much is due to the reduced frequency of incidents. Eliminating data related to incidents would directly yield the effect of control on flow rates. Accidents and vehicle stalls could then be evaluated separately, either by longer periods of record keeping or by drawing correlations between incidents and other parameters such as speed and the mean velocity gradient. Any increase in flow due to fewer stoppages would be more than the increase due to control.

The use of 30-sec flow rates facilitates the extraction of data during incidents and provides a sufficient number of points for the construction of flow relationships. Zero flow rates for a period of at least 2 or 3 min is enough to identify the time at which an incident took place. In this study, 30-sec flow data were printed on computer cards, and the appropriate cards were removed for final analysis. One- or 2-min intervals also would permit removal of data occurring during incidents.

Selection of Appropriate Control Strategy

The decision-maker is nearly always faced with the problem of making trade-offs. Analysis of traffic flow does not alleviate this dilemma. One cannot achieve both smoothest flow and maximum flow rates at the same time. The problem is somewhat easier to solve for tunnels because high rates of flow can be maintained at high speeds. It is not wise to strive for excessive speeds, however. One way to select the control strategies is to develop a speed-volume curve by using average speeds and flows at the bottleneck for each data collection period as data points. This was done for 19 days [4 days with no control and 15 days with 4 control strategies (Fig. 5)]. Because no off-peak flow data were obtained, all of the points are close to the peak of the curve. No regression analysis was performed on this curve. Each control strategy is represented by a different symbol so that a range of operation for each strategy can be identified. If a shorter cycle length had been attempted, more points would be on the high-speed side of the curve. The decision-maker may take a curve such as this and select whatever strategy may be appropriate for a given traffic condition. For instance, on a hot day when stalls are more likely to occur, the decision-maker may wish to minimize their occurrence within the tunnel and set the cycle for a high speed. He or she may do the same for Friday traffic if a high percentage of drivers not familiar with the tunnel is a problem. When the strategies have been selected and implemented, the traffic engineer may then proceed to monitor speeds, flow rates, and incident rates over a longer period of time to measure the full effect of the control system.

EFFECTIVENESS OF PRETIMED SIGNAL CONTROL

From the analysis of data obtained from the Baltimore Harbor Tunnel, we believe that pretimed signal control offers the tunnel administrator an effective and practical tool that can significantly improve the peak-period congestion problem. In addition, use of this inexpensive control system may lead to the use of more complex control measures.

Analysis of Control Alternatives at the Baltimore Harbor Tunnel

From Table 2, it is evident that metering produced marked effects on speeds at the bottleneck. As would be expected, the 120-sec cycle caused speeds to be highest, but the 160- and 180-sec cycles also caused substantial increases. The 240-sec cycle was essentially identical to an uncontrolled alternative.

An examination of the profiles of speed shown in Figures 2 and 6 indicates that a bottleneck does not exist at the foot of the upgrade when speeds are relatively high. Speeds tend to drop slightly as vehicles proceed along the upgrade probably because of the effect of both the 3.5 percent grade and the horizontal curve near the exit of the tunnel. Also speeds increase at the entrance portal of the tunnel; this indicates that a bottleneck exists somewhere upstream of that point. An extension of the profile to the signal would indicate that the control system now acts as the bottleneck because speeds at the signal seldom exceed 25 mph (40.2 km/h). This validates the primary purpose of control, which was that the bottleneck should be shifted to a point of higher capacity.

Flow rates for the control strategies also revealed a significant improvement over the no-control case. Weighted average hourly flows indicate that the 120- and 180-sec cycles were the most desirable. Indeed, other research and literature indicate that a more than 7 percent flow increase is substantial especially considering the simplicity of the system and that the entire increase is due solely to control. To the extent that incidents are reduced because of control, the increase will be even greater. The results of the 85 runs with the traffic analyzer indicate that a significant improvement in the quality of flow (indicative of safety) was realized for all but the 240-sec control alternative (Table 2). It is likely, then, that the number of incidents will be reduced for the controlled situation. It should be mentioned again that this increase is for the left lane only and is not an overall increase although similar increases are very likely possible in the right lane. Because some technical difficulties in right-lane hardware were encountered during data collection, only a limited amount of data was obtained. However, analysis of 1 day's data for each strategy in the right lane indicates that flow increases as high as 10 to 11 percent may be obtainable for both lanes combined.

Theoretically, the 160-sec cycle should have produced results somewhere between the 2- and 3-min cycles in traffic throughput, but this was not the case. This may have occurred because of a gap insufficient to allow the tunnel to be cleared of traffic before data collection was begun. The results of a t-test revealed that with a 95 percent level of confidence both the 2- and 3-min cycles were better than no control.

The speed-volume curve shown in Figure 5 does not decisively indicate the optimum cycle. The 2-min cycle has several advantages; the primary one is improved level of service. Therefore, if pretimed control were to be employed full-time at the Baltimore Harbor Tunnel, the 120-sec cycle should be used. An even shorter cycle for Fridays and the summer months might be considered.

If the capacity of the left lane of the tunnel for pretimed control is the apex of the speed-volume curve (1,530 vph), the control system has the potential to increase the capacity of that lane by 9 to 10 percent above normal operation. If the reduced frequency of incidents is taken into account, the potential increase is even greater, perhaps as high as 15 percent if an average of 1 incident per peak period is eliminated. This level of flow will be difficult to maintain in practice because of the unpredictable nature of driver behavior within the tunnel.

Guidelines for the Use of Control

The following procedures should be applied if the critical bottleneck has been identified as being at the foot of the upgrade:

1. Install speed and volume detectors at the bottleneck;
2. Determine average speeds and flow rates at the bottleneck under congested conditions;
3. Install signals with driver information systems at appropriate locations;

Figure 5. Speed-volume curve for 19 days of data collection.

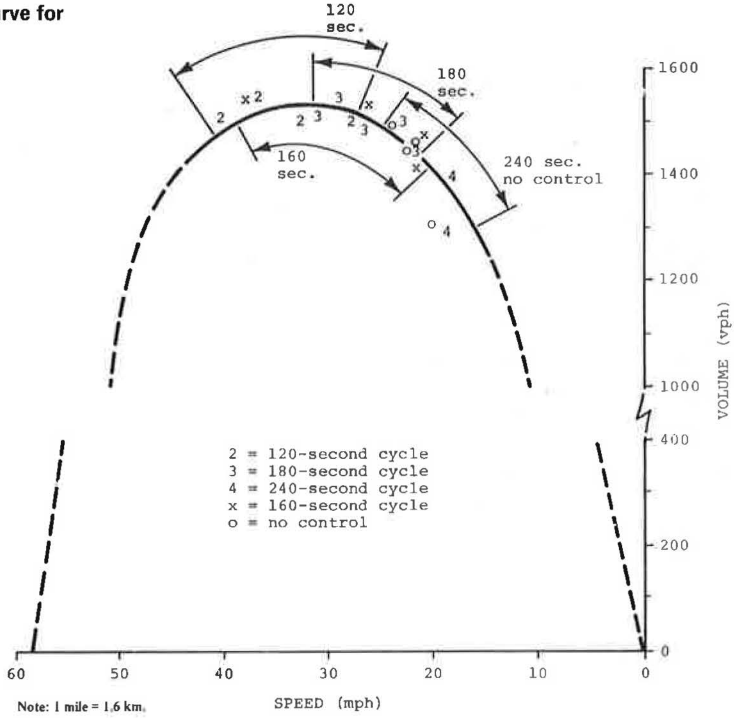


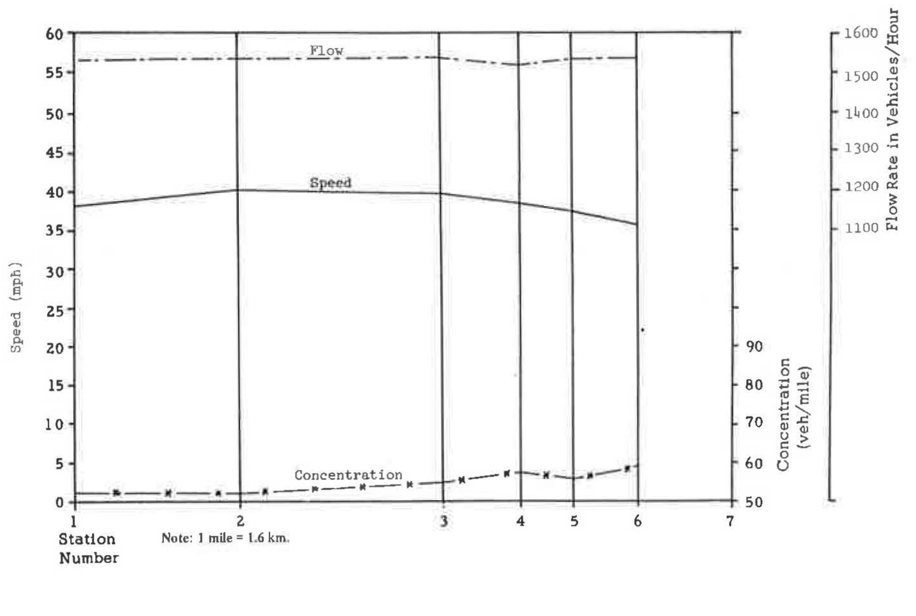
Table 2. Comparison of mean flow rates, mean speeds, and mean velocity gradients.

Strategy	Mean Speeds[a] (mph)	Increase Over Normal: Number	Increase Over Normal: Percent	Mean Hourly Flows[a] (vph)	Increase Over Normal: Number	Increase Over Normal: Percent	Mean Velocity Gradient[b] (sec^{-1})
No control	21.9	—	—	1,388	—	—	0.0327
120 sec	34.2	12.3	56.2	1,493	105	7.6	0.0218
160 sec	26.6	4.7	21.5	1,469	81	5.8	0.0248
180 sec	26.1	4.2	19.2	1,495	107	7.7	0.0256
240 sec	19.6	-2.3	-10.5	1,366	-22	-1.6	0.0367

Note: 1 mile = 1.6 km.

[a]Measured at bottleneck in left lane. [b]Measured through entire tunnel in both lanes.

Figure 6. Traffic-flow profiles, Feb. 6, 1973, left lane, 120-sec cycle.



4. Determine capacity at the signal for various cycle lengths and splits;
5. Select several cycle lengths; some should have capacities higher than those of the bottleneck and some should have capacities lower than those of the bottleneck;
6. Test each alternative strategy for several days, and make cycle adjustments if necessary;
7. Evaluate the tested strategies and select those that yield the best combination of flow rates and quality of flow; and
8. Implement chosen strategies and continue long-term monitoring of flows and incident rates.

The proposed metering system is rather primitive when it is compared to some of the extensive control systems being developed. But a solution need not be complex to be practical. Pretimed signal control may be the tool that can fit both the needs and capabilities of a staff managing a tunnel facility. Furthermore, if simple control systems can build the confidence of those authorized to spend the money for their use, the path will be cleared for more widespread use of sophisticated control measures. The initial outlay of funds for a pretimed system can be more easily rationalized if there is little to lose and much to gain.

CONCLUSIONS

Experimentation with and evaluation of the pretimed traffic-control system on the Baltimore Harbor Tunnel has yielded a number of conclusions.

A pretimed signal-control system has several advantages.

1. The system can increase flow rates. Left-lane flows in the Harbor Tunnel were increased by about 7 percent at the bottleneck.
2. The system can improve the quality of flow through the entire tunnel. Speeds in the Baltimore Harbor Tunnel for the 120-sec cycle were 56 percent higher than were speeds with no control.
3. Equipment is low in cost and easy to maintain.
4. Cycle lengths and splits may be designed for different traffic situations. The 120-sec cycle provided a good balance of flow rates and flow quality in the Baltimore Harbor Tunnel.
5. Equipment may be gradually added to the basic system to further improve traffic flow.
6. The system can be used to improve flows both for peak hours and for clearing traffic rapidly after incidents.

The disadvantage of the pretimed signal-control system is that it cannot react automatically to traffic conditions within the tunnel unless it is supplemented with computerized detectors.

Speed data at the foot of the upgrade indicated that the Baltimore Harbor Tunnel operates essentially as a 2-state system at either high or low speeds with few speeds in between.

Elimination of data during the occurrence of incidents is a good means to evaluate the effects of control alone. Incidents then can be examined independently.

Pretimed control may be a stepping-stone to the widespread acceptance of more sophisticated control measures.

RECOMMENDATIONS

Data from other tunnels should be collected for the purposes of

1. Determining whether the framework presented here offers a solution to the problems in tunnels other than the one studied here;

2. Supporting the hypothesis of 2-state tunnel operation;
3. Increasing existing knowledge of tunnel traffic flow; and
4. Examining how flows in the right lane are affected by control.

Methods of sensing traffic characteristics within tunnels and methods of recording them more accurately without affecting traffic flow must be determined.

The apparent discontinuity in the q-k curve for bottlenecks should be examined further.

ACKNOWLEDGMENT

The opinions, findings, and conclusions expressed in this report are those of the authors, and not necessarily those of the Maryland Highway Administration or the Federal Highway Administration.

REFERENCES

1. R. S. Foote. Vehicle Tunnel Surveillance to Improve the Level of Traffic Service. Port of New York Authority, 1964.
2. E. C. Carter and S. P. Palaniswamy. Study of Traffic Flow on a Restricted Facility, Interim Report—Phase I. Dept. of Civil Engineering, Univ. of Maryland, 1973.
3. P. F. Everall. Urban Freeway Surveillance and Control: The State-of-the-Art. Federal Highway Administration, 1972.
4. S. A. Smith and E. C. Carter. Pre-Time Signal Control System for a Vehicular Tunnel, Interim Report II-3. Dept. of Civil Engineering, Univ. of Maryland, July 1974.
5. R. S. Foote, K. W. Crowley, and A. T. Gonseth. Tunnel Traffic Capacity Study—Report VII. Port of New York Authority, 1960.
6. R. S. Foote. Tunnel Traffic Research. Port of New York Authority, 1968.
7. S. P. Palaniswamy. Optimization of Traffic Flow on a Restricted Facility—A Conceptual Framework. Dept. of Civil Engineering, Univ. of Maryland, PhD dissertation, 1973.
8. M. R. Palmer. The Development of Traffic Congestion. Bureau of Highway Traffic, Yale Univ., 1959.
9. W. Helly and P. C. Baker. Acceleration Noise in a Congested Signalized Environment. In Vehicular Traffic Science (L. C. Edie, R. Herman, and R. Rothery, eds.), American Elsevier Publishing Co., New York, 1967.
10. Highway Capacity Manual. HRB Spec. Rept. 87, 1965, 411 pp.

DETECTING FREEWAY INCIDENTS UNDER LOW-VOLUME CONDITIONS

Conrad L. Dudek, Graeme D. Weaver, Gene P. Ritch, and Carroll J. Messer,
Texas Transportation Institute, Texas A&M University

Two computer algorithms for automatic, freeway-incident detection under low-volume conditions were developed. The first approach uses a time-scan process. The second approach, considered to be superior to the first, operates on an event-scan principle. Computer simulations produced a family of curves that are useful in determining sensor-spacing requirements for an operational system using the event-scan algorithm. The results indicate that, when detector spacings of 1,000 ft (304.8 m) are used, all incidents on a 3-lane freeway section can be detected within 3 min for volumes up to 500 vehicles per hour. When volumes approach 1,000 vehicles per hour, 85 percent of the incidents can be detected within 3 min. Faster detection capabilities at the higher volumes would require closer detector spacings. Incident-detection operational considerations, particularly the manner in which software can be developed to recognize and compensate for vehicle-count errors produced by semitrailers and lane-change maneuvers, also are discussed.

•RECENT research has focused on the development of freeway-incident-detection algorithms for effective traffic management within freeway corridors. Work, notably that by Courage and Levin (1), Schaefer (2), Cook and Cleveland (3), Dudek and Messer (4), and Dudek, Messer, and Nuckles (5), has been directed primarily toward detecting incidents under medium- and heavy-flow conditions.

Although peak-period operation rightly commands most of the attention in freeway operations, the freeway operates 24 hours a day, and, during about 20 hours each day, most freeways operate below peak-volume conditions. However, certain safety problems continue to exist. One such problem is the accident or disabled vehicle on or adjacent to a main lane that, when approached by an unsuspecting driver at high speed, provides potential for a severe collision or at least a sudden change in the operating characteristics of the approaching vehicle. This problem is even more severe in freeway sections where sight distance is restricted by geometric features such as overpasses or horizontal curves coupled with median fences and retaining walls. In addition, freeway drivers operating under light-flow and high-speed conditions expect that the road ahead will be free of restriction; thus an unexpected event such as a stopped vehicle can create a greater hazard under these conditions than under alerted conditions. The problem is compounded when the incident occurs on elevated freeway sections, causeways, and tunnels. Two methods for detecting vehicular incidents under low-volume conditions are presented in this paper.

APPROACH

Control Variables

Incident detection under low-volume conditions requires a different basic approach than that used for high-volume situations. Basically, incident-detection algorithms for heavy-flow conditions rely on the measurements of flow discontinuities resulting from the reduced capacity created by the incident. During light-flow conditions, stoppage waves will not readily propagate (5). In selecting the control variable for incident detection under light-volume conditions, several variables used in incident-detection algorithms for peak periods are therefore unsatisfactory.

Because speeds are high and fairly uniform along each segment of the freeway when volumes are extremely light, the use of vehicle storage concepts appears to represent a favorable control method. Total input-output analysis appears to be unsatisfactory, however, because under light flow, which allows ample maneuvering space and potential for high-speed passing, a vehicle conceivably can enter the control section at a very high rate of speed, overtake a slower vehicle in the control section, and actually emerge from the section before the slower vehicle. Therefore, the speed variable should be considered in addition to the number of vehicles within the control section at any time. To accomplish this, the input-output technique was refined from total input-output analysis to individual-vehicle input-output analysis based on the time and speed entering the control section and time of exit from the control section as determined by a computer. The expected exit time would be

$$t_e = t_i + \frac{D}{V}$$

where

t_e = exit time in seconds,
t_i = entrance time in seconds,
D = distance between detectors in feet (meters), and
V = speed of vehicle in feet per second (meters per second).

This relationship is based on the assumption that vehicle speed remains constant between detectors.

Under this concept, the control variables are speed and the time that a vehicle enters and leaves the system. One can measure these variables by using lane detectors in pattern arrangements that are now used in many freeway-control systems.

Incident-Detection Algorithms

System A—Time-Scan Operation

Vehicle accounting in this system is accomplished on a fixed time interval. The relationship of time interval and detector spacing becomes quite critical to the expected results. Obviously, it would be desirable to place detectors at very short intervals throughout the control section. This would permit almost continuous monitoring of speed and vehicle count throughout the section with very small speed changes between consecutive detectors. Economics, however, prohibit such a luxury. Therefore, it becomes necessary to try to optimize the interval of detector spacing and accounting time to arrive at a compromise that is tolerable from both an economic standpoint and a false-alarm rate.

Figure 1 illustrates the operation of system A. Suppose that the system is turned on at T_0. All detectors are awaiting a vehicle actuation from which time of activation and vehicle speed can be recorded. At entrance time TA1, a vehicle crosses the detectors at location A and is registered in the system. Vehicle speed is measured, and predicted exit time, TB1, is computed. The slope of the line between TA1 and TB1 represents the speed. Another vehicle enters at TA2 traveling at a high rate of speed, and its exit time, TB2, is computed. As shown in Figure 1, this vehicle was traveling fast enough to pass the first vehicle in the system and would be expected to exit at location B before the first vehicle. Other vehicles entered the segment A-B as shown, and

Figure 1. System A, time-scan operation.

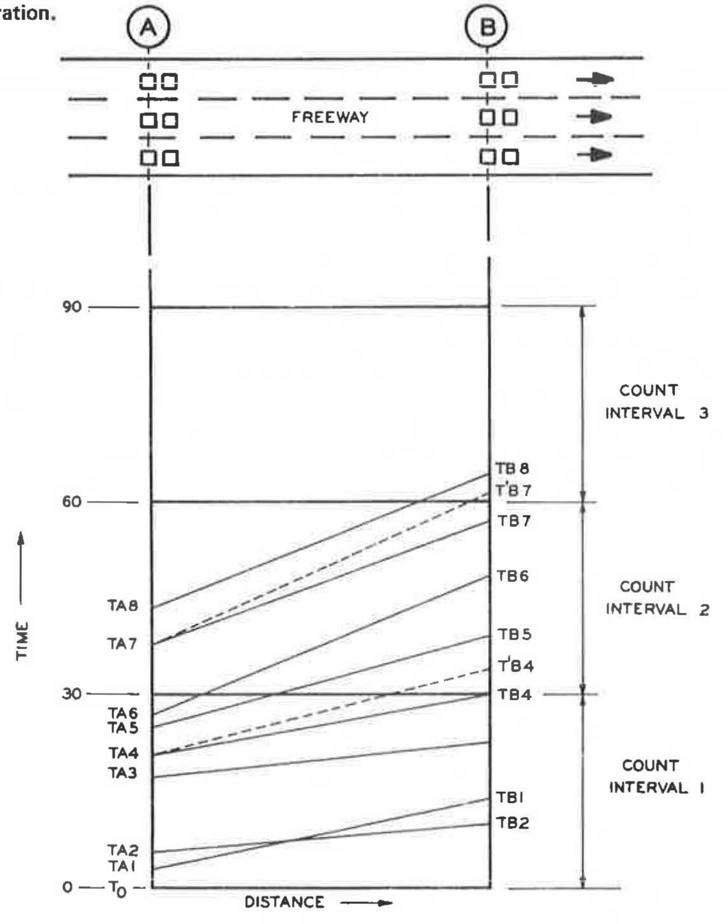
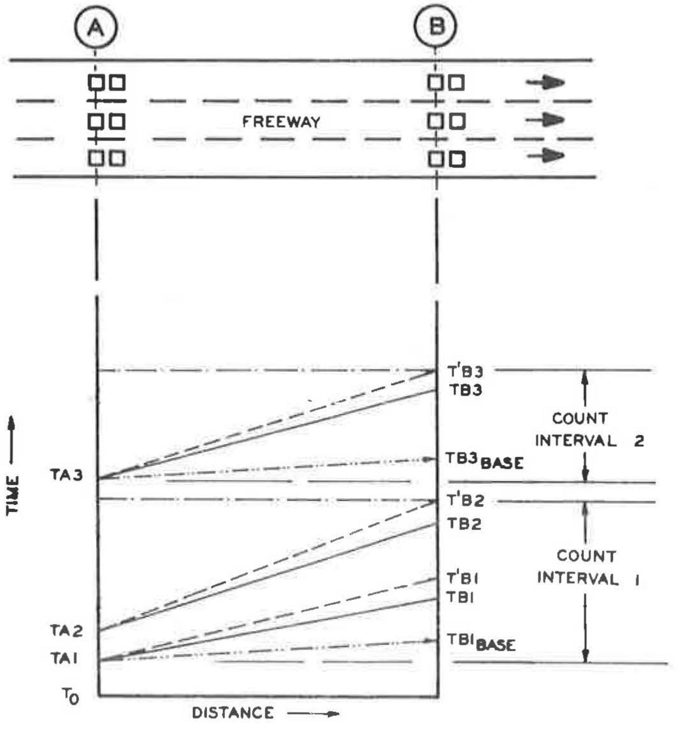


Figure 2. System B, event-scan operation.



expected exit times at B were computed.

Using a 30-s accounting interval, as shown in Figure 1, one can see that, within the first time interval, 6 vehicles entered at location A, and, within that same time interval, 4 would be expected to exit at B. The other 2 vehicles (vehicles 5 and 6) would be expected to arrive at B during the next 30-s interval, and they will be considered during that time interval. Expected exits are compared to actual exits at B. If the number exiting is less than expected, the assumption would be made that 1 or more vehicles, depending on the disparity, stopped between locations A and B. Theoretically, this logic appears to be valid. In practice, however, speeds vary, but the model assumes constant speed between detectors.

Consider vehicle 4 as an example of potential false alarm. Vehicle 4 was expected to arrive at B at TB = 30 s. Had the speed measurement been slightly high, or had vehicle 4 reduced speed slightly after passing the detectors at A, the actual arrival time at B would be slightly greater than the 30-s time point as shown by the dashed line in Figure 1. Thus, although it arrived without incident, a false alarm would have occurred. Detection delays could be incorporated to check the next accounting interval to determine whether the number of exiting vehicles was 1 more than expected to balance the system. However, the possibility does exist that the same phenomenon could occur during the next time interval. Therefore, the first late-arrival actuation would cancel out the second expected actuation and thus cause a false alarm. It becomes apparent that the false-alarm rate will increase as volumes and detector spacings increase.

System B—Event-Scan Operation

To alleviate the potential false-alarm problems associated with the time-scan operation, a variable time interval for vehicle accounting was developed.

Figure 2 illustrates the operation of system B. Assume that the system is turned on at T_0. When the first vehicle arrives at location A at TA1, 3 computations are made. First, the base time at B is computed. This represents the shortest practical time that the vehicle could be expected to arrive at detector B. Assuming a maximum speed of 100 mph (160.9 km/h) would be feasible because few vehicles could be expected to exceed this speed. Second, the expected arrival time at B, TB1, is computed based on measured speed. Third, to compensate for errors in speed measurements from the detectors, an allowable, speed-reduction safety factor of 10 percent is applied to the measured speed and a late expected arrival at B, T'B1, is computed. From this, a time is projected back to location A that is a base time determined from the assumed base speed. If a second actuation did not occur at A before this projected base time, the accounting interval would be established at A as the interval between TA1 and $TA1_{Base}$ and at B as the interval between $TB1_{Base}$ and TB1. As shown in Figure 2, a second vehicle did arrive at A before the projected base time. Therefore, the process is repeated. The example in Figure 2 shows that a third vehicle did not arrive between TA2 and $TA2_{Base}$; therefore, the time intervals at A and B are established as indicated. The sequence begins again when vehicle 3 crosses the detectors at A. The second time interval shown in Figure 2 indicates that only 1 vehicle arrived in the interval. The time interval under system B will differ in length according to the arrival rate at the first set of detectors. In practice each consecutive pair of detector sets would constitute a subsystem, and the accounting process would be accomplished throughout each subsystem whenever the vehicles cleared each subsystem.

It can be seen that, as flow rates increase, extension of the time interval can be expected and may become so long that insufficient response time could be provided after vehicle accounting procedures. The simulation studies discussed in the next section identify probabilities of detection by using the event-scan algorithm for various sensor spacings.

DETECTOR SPACING REQUIREMENTS FOR EVENT-SCAN OPERATION

This section discusses the influence of detector spacing on automatic incident detection for low-volume conditions. Only the event-scan operation (algorithm B) is discussed because it is considered to be the superior approach.

A computer simulation program was developed and run for a 3-lane directional freeway section. Volumes of 100, 500, and 1,000 vehicles per hour (vph) were tested with detector spacings of 500, 1,000, and 1,500 ft (152.4, 304.8, and 457.2 m) respectively. Ten hours of simulated traffic flow were produced for each of the 9 combinations of volumes and sensor spacings. The program was developed with the assumption that each vehicle entering the system had an equal probability of becoming disabled or involved in an incident. Poisson arrivals were assumed, and speed distributions collected on each lane of the Gulf Freeway in Houston, Texas, for the selected volumes were incorporated into the program.

The percentage of incidents detected within given time periods based on the simulation results are shown in Figures 3, 4, and 5. The average, smallest, and largest detection times for each volume and detector spacing combination are given in Table 1. The results illustrate incident-detection capabilities under low-volume conditions for the event-scan operation.

The results indicate that for volumes of 500 vph or less detector spacings of 1,000 ft (304.8 m) would provide adequate incident-detection response on a 3-lane freeway section. At volumes of 1,000 vph and greater, detector spacings of less than 1,000 ft (304.8 m) should be considered.

SOME OPERATIONAL CONSIDERATIONS

Incident detection under low-volume conditions that uses the algorithms previously described places stringent requirements on the surveillance system that have not been necessary for other freeway-operational-control functions. These include ramp metering, shock-wave detection, and incident detection during medium- and high-volume conditions. Accurate vehicle counts and relatively accurate speed measurements are essential if the system is to operate effectively. Experiences on the Gulf Freeway surveillance and control systems have indicated that it is not always possible to obtain perfectly accurate vehicle counts. Preliminary studies have produced an average of 1 error per 10 min at 200 vph on the 3-lane directional freeway. For automatic incident detection, this would mean a false alarm or a failure to detect an incident. Therefore, special studies were conducted during light-flow conditions to determine the source of error and to develop methods to compensate for it. Data from detectors on the Gulf Freeway were automatically processed by the computer system and produced printout data in real time. Observers using television monitors viewed traffic passing across the detectors and noted any irregularities between the traffic and the computer output.

Equipment

Traffic data were collected from 1 of the many sets of loop detectors on the inbound Gulf Freeway. Two loop detectors are positioned in each of the 3 lanes. Each loop detector is composed of 3 coils of 14-gauge wire installed in a saw cut 6 ft (1.83 m) square centered in the 12-ft (3.66-m) lane. The leading edges of the lead and lag loops are separated by 18 ft (5.49 m).

Program

The data acquisition programs within the digital computer operate under a real-time,

Figure 3. Incident-detection algorithm performance for detector spacings of 500 ft (152.4 m).

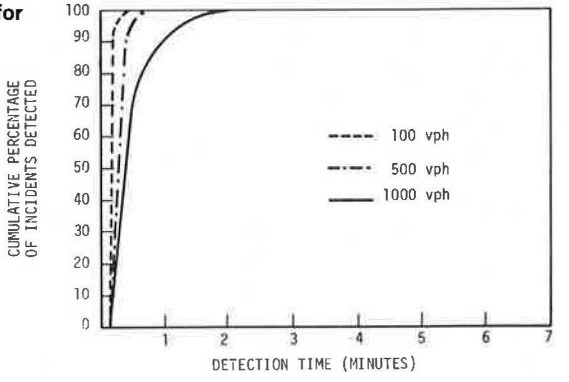


Figure 4. Incident-detection algorithm performance for detector spacings of 1,000 ft (304.8 m).

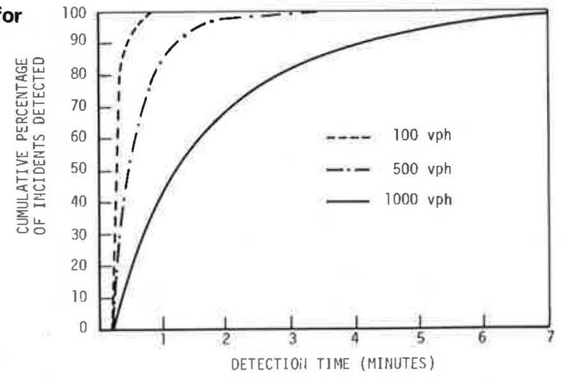


Figure 5. Incident-detection algorithm performance for detector spacings of 1,500 ft (457.2 m).

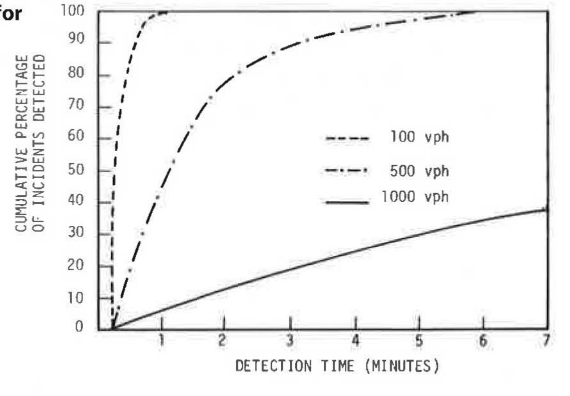


Table 1. Incident-detection times.

	Detection Time (s)		
Spacing Combination	Average	Smallest	Largest
Detectors 500 ft apart			
100 vph	7	4	20
500 vph	7	4	40
1,000 vph	13	4	133
Detectors 1,000 ft apart			
100 vph	15	9	50
500 vph	25	9	216
1,000 vph	74	9	518
Detectors 1,500 ft apart			
100 vph	24	13	80
500 vph	62	13	427
1,000 vph	590	15	3,000

Note: 1 ft = 0.3048 m.

multiprocessing, operating system. For each loop-detector-relay operation, programs are initiated that establish the time a vehicle enters or leaves the field of influence. Program timings are accurate to the nearest millisecond.

Table 2 gives the output format of the program timings. The clock time in column 1 in milliseconds is cyclic in nature; it increments from 0 to 32 575 ms and then decrements from 32 576 to 0 ms. It is used to establish an event occurrence. In this case, it is the time when a vehicle enters or leaves a loop detector.

Columns 2 through 7 indicate the headways in milliseconds between operations of lead and lag detectors for the inside, middle, and outside lanes. These values are significant when one determines whether more than 1 detector activation occurred for a given vehicle. If a headway of less than 500 ms is found, detector multiactivation or malfunction is suspected and should be further analyzed.

Columns 8 through 13 indicate the passage time or occupancy in milliseconds of a vehicle over a loop detector. These timed values are useful as vehicle "signatures" for identification purposes.

Volumes, headways, and occupancies registered by the lead and lag detectors can be compared on each lane to determine whether discrepancies exist. For example, at 374 ms in Table 2, the system indicates that a vehicle arrived at the lead detector on the middle lane 3651 ms after the previous vehicle. At 775 ms, the vehicle occupancy over the lead detector is 331 ms. Similarly, the headway and vehicle occupancy over the lag detector are 3731 and 305 ms. Comparison of the successive data in Table 2 reveals consistent results between the lead and lag detectors.

Columns 14 through 16 indicate the individual vehicle speeds in miles (kilometers) per hour in each lane. They are calculated values. The travel time between the lead and lag detectors is divided into an effective distance although the actual distance is 18 ft (5.49 m). The distance between loop detectors had to be changed to some value other than 18 ft (5.49 m). This was based on previous research work on the Gulf Freeway. A loop-tuning program in the computer was executed during free-flow traffic conditions. The effective distance was changed so that the calculated speed fell within the known free-flow traffic speed [48 to 56 mph (77.2 to 90.1 km/h)]. Unreasonable speed values in columns 14, 15, and 16 may exist because loop-detector amplifiers were changed and the effective distances were not recalibrated. The actual effective distances used for all data samples included in this report were 18.44, 19.55, and 19.65 ft (5.62, 5.96, and 5.99 m) for the inside, middle, and outside lanes respectively.

Using the above data acquisition program with real-time printout capabilities, we were able to observe traffic flow by closed-circuit television and compare the observed events to the data output.

Trucks

Table 2 serves to illustrate the normal actuation of semitrailers. Note the large occupancy values in the middle lane (columns 9 and 12) at 6260 and 6488 ms.

On many occasions, large trucks will cause double actuations by the lead or lag detectors or both. The data given in Table 3 indicate that a semitrailer caused the lead detector to register 1 actuation and caused the lag detector to register 2 actuations. Scrutiny of the headway and speeds provides a means for correcting for the double actuation. Although occupancies of 306 and 196 ms are within the range of acceptable data, the headway resulting from the lag detector registering the trailer as a second vehicle was computed at 441 ms. This is below normal expectations. Also the speed of 23 mph (37.0 km/h) reinforces the fact that a double actuation on the lag detector occurred. This information will allow development of software to compensate for double actuations.

Lane Changing

Another factor affecting the accuracy of the input or output vehicle count is lane

Table 2. Data acquisition program output.

Clock Time (ms)	Headway (ms)						Occupancy (ms)						Speed (mph)		
	Lead			Lag			Lead			Lag					
(1)	Inside (2)	Middle (3)	Outside (4)	Inside (5)	Middle (6)	Outside (7)	Inside (8)	Middle (9)	Outside (10)	Inside (11)	Middle (12)	Outside (13)	Inside (14)	Middle (15)	Outside (16)
0	0	0	0	0	0	1083	0	0	317	0	0	0	57	59	53
374	0	3651	0	0	0	0	0	0	0	278	0	268	57	59	53
775	0	0	0	0	3731	0	0	331	0	0	305	0	57	51	53
1273	1426	0	0	1416	0	0	247	0	0	0	0	0	59	51	53
1572	0	1219	0	0	0	0	0	0	0	244	0	0	59	51	53
1972	0	0	0	0	1188	0	0	251	0	0	233	0	59	56	53
3238	0	1577	0	0	1553	0	0	296	0	0	0	0	59	52	53
3473	0	0	0	0	0	0	0	0	0	0	275	0	59	52	53
4546	0	1332	0	0	1324	0	0	273	0	0	0	0	59	54	53
4772	0	0	0	0	0	0	0	0	0	0	250	0	59	54	53
5662	4305	1356	0	4321	0	0	331	0	0	0	0	0	55	54	53
5856	0	0	0	0	0	0	0	0	0	306	0	0	55	54	53
6260	0	0	0	0	1345	0	0	630	0	0	0	0	55	57	53
6488	0	0	0	0	0	0	0	0	0	0	621	0	55	57	53

Note: 1 mile = 1.61 km.

Table 3. Double actuation from semitrailer.

Clock Time (ms)	Headway (ms)						Occupancy (ms)						Speed (mph)		
	Lead			Lag			Lead			Lag					
(1)	Inside (2)	Middle (3)	Outside (4)	Inside (5)	Middle (6)	Outside (7)	Inside (8)	Middle (9)	Outside (10)	Inside (11)	Middle (12)	Outside (13)	Inside (14)	Middle (15)	Outside (16)
22 844	0	0	0	2 418	0	0	279	0	0	0	0	0	59	58	59
22 647	0	0	0	0	0	0	0	0	0	272	0	0	59	58	59
12 017	10 740	0	0	10 750	0	0	276	0	0	0	0	0	56	58	59
11 909	0	0	0	0	0	0	0	0	0	255	0	0	56	58	59
9 848	0	12 959	0	0	12 969	0	0	0	0	0	306	0	56	58	59
9 743	0	0	0	0	0	0	0	647	0	0	0	0	56	58	59
9 511	2 783	0	0	0	441	0	0	0	0	0	196	0	56	23	59
9 319	0	0	0	2 776	0	0	280	0	0	0	0	0	58	23	59
9 050	0	0	0	0	0	0	0	0	0	268	0	0	58	23	59
8 491	0	1 625	0	0	1 188	0	0	273	0	0	0	0	58	56	59

Note: 1 mile = 1.61 km.

Table 4. Double actuations from lane change and detector amplifier differences.

Clock Time (ms)	Headway (ms)						Occupancy (ms)						Speed (mph)		
	Lead			Lag			Lead			Lag					
(1)	Inside (2)	Middle (3)	Outside (4)	Inside (5)	Middle (6)	Outside (7)	Inside (8)	Middle (9)	Outside (10)	Inside (11)	Middle (12)	Outside (13)	Inside (14)	Middle (15)	Outside (16)
17 965	0	0	0	0	0	5 848	0	0	0	0	0	297	59	61	55
17 756	0	0	0	0	0	0	0	0	256	0	0	0	59	61	55
15 978	7 767	0	2 067	0	0	0	0	0	0	0	0	0	59	61	55
15 827	0	0	0	7 779	2 393	2 098	317	0	0	0	0	163	55	06	49
15 792	0	0	0	0	0	0	0	0	0	0	0	0	55	06	49
15 573	0	0	0	0	0	0	0	0	328	299	58	0	55	06	49
12 392	0	0	3 518	0	0	3 482	0	0	0	0	0	298	55	06	56
12 175	0	0	0	0	0	0	0	0	0	0	0	0	55	06	56
10 636	0	7 647	0	0	0	0	0	168	0	0	0	0	55	06	56
10 399	0	0	0	0	5 271	0	0	0	0	0	162	0	55	56	56
9 829	0	834	0	0	0	0	0	140	0	0	0	0	55	56	56
9 600	0	0	0	0	829	0	0	0	0	0	132	0	55	57	56

Note: 1 mile = 1.61 km.

Table 5. Low occupancy resulting from a motorcycle.

Clock Time (ms)	Headway (ms)						Occupancy (ms)						Speed (mph)		
	Lead			Lag			Lead			Lag					
(1)	Inside (2)	Middle (3)	Outside (4)	Inside (5)	Middle (6)	Outside (7)	Inside (8)	Middle (9)	Outside (10)	Inside (11)	Middle (12)	Outside (13)	Inside (14)	Middle (15)	Outside (16)
18 820	0	0	4 226	0	0	4 197	0	0	330	0	0	0	51	59	56
19 013	0	0	0	0	0	0	0	0	0	0	0	275	51	59	56
19 350	0	4 800	0	0	0	0	0	56	0	0	0	0	51	59	56
19 588	0	0	0	0	4 818	0	0	0	0	0	51	0	51	54	56
20 408	3 825	980	0	3 799	0	0	290	0	0	0	0	0	57	54	56
20 539	0	0	0	0	977	0	0	270	0	0	0	0	57	55	56
20 624	0	0	0	0	0	0	0	0	0	295	0	0	57	55	56
20 818	0	0	0	0	0	0	0	0	0	0	249	0	57	55	56

Note: 1 mile = 1.61 km.

changing in the vicinity of the sensors. The resultant characteristics of lane changes are given in Table 4. A vehicle traveling in the inside lane changed to the middle lane and entered the area of the middle-lane lag detector. This caused a vehicle occupancy computation on the middle-lane lag detector without a corresponding activation on the middle-lane lead detector. Note also an extremely low speed value on the middle lane. These patterns can be readily recognized by the computer to automatically adjust for discrepancies, which ensures a higher degree of accuracy at the input and output count stations.

Detector Amplifier Differences

Table 4 illustrates differences in amplifier measurements. Amplifiers produced by a different manufacturer than the one that produced those used on the inside and outside lanes were placed in the middle lane. The traffic moved over all lanes at free-flow speed, yet the middle lane (columns 9 and 12) clearly indicates much smaller travel times across the lead and lag loops. The detection operations are usable except when very small values (50 to 100 ms) occur. The existence of a vehicle or the occurrence of a lane change poses problems that must be solved. Using an amplifier with this characteristic could make the establishment of data limits very difficult.

Motorcycles

Although detector occupancies of less than 100 ms generally are suspect, motorcycles produce low values (some detector amplifiers will not even detect motorcycles). For example, a motorcycle traveling on the middle lane had occupancies of 56 and 51 ms registered by the lead and lag detectors respectively (Table 5). Examination of the headways, speeds, and number of activations reveals no unusual patterns. Thus the system could be designed to recognize the presence of motorcycles.

SUMMARY

Two computer algorithms for automatic freeway-incident detection under low-volume conditions were developed and presented. Both approaches used input-output techniques that require accurate vehicle counts. Vehicle speeds were computed at the input station, and times of departures at the output stations were determined. One approach used a time-scan process. The second, considered to be the superior of the 2 approaches, operated on an event-scan principle.

The computer simulations produced a family of curves that are useful in determining sensor-spacing requirements for an operational system that uses the event-scan algorithm. The results indicate that 1,000-ft (304.8-m) detector spacings will provide adequate response to incidents for volumes up to 500 vph on a 3-lane freeway section. At volumes of 1,000 vph and greater, detector spacings of less than 1,000 ft (304.8 m) should be considered.

Because accurate counts are essential at both the input and output sensor stations, the study results have shown that volume counts must be supplemented by pattern recognition of headway, vehicle occupancy, and speed data to compensate for volume errors produced by semitrailers and lane-change maneuvers.

ACKNOWLEDGMENTS

This paper discusses one phase of the research project, Development of Urban Traffic Management and Control Systems, conducted by the Texas Transportation Institute and the Texas Highway Department in cooperation with the Federal Highway Administration. The contents of this paper reflect the views of the authors who are responsible for the

facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Federal Highway Administration.

REFERENCES

1. K. G. Courage and M. Levin. A Freeway Corridor Surveillance, Information, and Control System. Texas Transportation Institute, Texas A&M Univ., Research Rept. 488-8, Dec. 1968.
2. W. E. Schaefer. California Freeway Surveillance System. California Division of Highways—District 7, Nov. 1969.
3. A. R. Cook and D. E. Cleveland. Detection of Freeway Capacity-Reducing Incidents by Traffic Stream Measurements. Transportation Research Record 495, 1974, pp. 1-11.
4. C. L. Dudek and C. J. Messer. Detecting Stoppage Waves for Freeway Control. Highway Research Record 469, 1973, pp. 1-14.
5. C. L. Dudek, C. J. Messer, and N. B. Nuckles. Incident Detection on Urban Freeways. Transportation Research Record 495, 1974, pp. 12-24.

DISCUSSION

Joseph A. Wattleworth, Department of Civil Engineering, University of Florida

Dudek et al. have started research efforts into a new and very difficult area: incident detection under conditions of low freeway volumes. Most incident detection has been incorporated into peak-period freeway-surveillance and freeway-control systems and has used large changes (or discontinuities) in 1 or more of the macroscopic flow variables (flow, speed, and density). Past research in incident detection has concentrated on the high-volume area because of the overall concern for peak-period problems and because incident detection is easier to accomplish during peak periods.

As Dudek et al. point out, incident detection is more difficult during low-volume conditions. Discontinuities in the traffic stream are smaller than they are under heavy-flow conditions and shock waves generally are not propagated readily under low-volume conditions. The traditional approaches to incident detection have been techniques in which either the affected motorist or a passing motorist reports the incident to the appropriate authorities. In some areas call boxes have been installed to facilitate the reporting of incidents; patrolling vehicles also have been used for this purpose. Little has been done to apply electronic-detection principles to low-volume incident detection.

Dudek et al. have made a pioneering effort into the low-volume incident-detection problem. They have investigated 2 low-volume incident-detection schemes: time-scan methods and event-scan methods. They have recommended the event-scan techniques. Basically, a set of freeway detectors senses each vehicle that crosses it and predicts its arrival time at the next detector station. When a vehicle does not arrive at the downstream detector station as predicted, the incident-detection logic classifies the event as an incident. Thus, the incident-detection system relies on an individual-vehicle-accounting system.

I have no questions about the theoretical validity of this approach, but I would like to raise the question of practicability based on current detector technology. It is generally accepted that a detector station will not provide a perfect total volume count. The count will be either high or low, and the detector configuration can be changed to make the detected count higher or lower, but errors are not eliminated. In other words, if the configuration or sensitivity is changed to reduce overcounting errors, undercounting errors will be increased. In any case, it does not appear at this time that there is a counting station that will yield perfect counts.

Any incident-detection system must decide between 2 types of errors:

1. Missing a real incident and
2. Identifying a nonincident as an incident.

The latter is a false alarm. If we assume an error rate percentage, E, an hourly volume, V in hundred vehicles per hour, and N detector stations per mile (kilometer), we would expect E × V × N false alarms per hour per mile (kilometer). For example, if we assume an average detector station error rate of 1 percent, a volume of 200 vph and 5 detector stations per mile (3 detector stations per kilometer), we would expect 10 false alarms per hour per mile (6 false alarms per hour per kilometer). Thus, even though there is a relatively low error rate and a low volume, a high generation rate of false alarms is experienced.

It would appear that the problem of low-volume incident detection should be subjected to a traditional systems analysis in which all reasonable alternatives would be explored. It would seem that other alternative schemes might be possible.

The detection scheme called for by the authors involves a pair of detectors in each lane of the freeway at each detector station, and the recommended spacing of the detector stations is every 1,000 ft (30.48 m). This would require about 32 detectors per mile (20 detectors per kilometer) for a 3-lane freeway section and 42 detectors per mile (26 detectors per kilometer) for a 4-lane freeway. Alternatively, these same detectors could be located on the shoulders of the freeway about every 200 ft (60.96 m). This system might provide a higher level of accuracy in detecting incidents that used the shoulders and almost certainly would have a much lower false-alarm rate.

Another possibility appears to be worth pursuing. Rather than use an input-output accounting procedure with its proven error problems, it might be better to look into the use of a detection system that would directly detect a stopped vehicle. The air traffic control radar system at the Tampa International Airport shows the movement of vehicles across the Howard Franklin Bridge. It probably is not able to discriminate between 2 vehicles in a platoon, but it most likely would be able to discriminate between stopped and moving vehicles, and it could identify stoppages. Further refinements of this area-detection system might produce a usable technique for detecting incidents under low-volume conditions.

In conclusion, Dudek et al. have made a worthwhile start into the area of low-volume incident detection. There is clearly much more work to be done, and the authors are encouraged to continue their efforts.

Joseph M. McDermott, Chicago Area Expressway Surveillance Project, Illinois Department of Transportation

The authors have proposed 2 algorithms for detecting freeway incidents under low-volume conditions. These algorithms may have merit for tunnels, bridges, and freeway sections without shoulders, but their application elsewhere requires considerable refinement. They need to distinguish critical from noncritical stoppages and integrate detection and response into a cost-effective system.

All incidents will be detected with no false alarms if the proposed detection system can be tuned and operated perfectly (a practically impossible and unrealistic event), if logic refinements can be incorporated to handle normal exit and entrance ramp changes along the freeway, and if lane changing at the detection points can be accounted for. The detected incidents, however, will include all vehicle stoppages along the freeway, including those made on the shoulders.

Numerous studies of stoppages and freeway shoulder use, including urban studies in Chicago, Detroit, and Houston, have shown that vehicle stoppages occur quite frequently, are usually on shoulders, and are of short duration; they usually do not involve disabled vehicles and usually require no assistance (6, 7, 8, 9, 10). For example, one 3-day Chicago-area study on an urban freeway section with 120,000 average daily traffic reported 1 vehicle stoppage for reasons other than congested traffic for every 2,500 vehicle miles (4023 vehicle kilometers) of travel, an average of 1 stopped vehicle per directional mile per hour (1.6 stopped vehicles per directional kilometer per hour) (7). The right shoulder handled 84.1 percent of the stopped vehicles, the left shoulder handled 7.4 percent. The main lanes accounted for 4.5 percent. The remaining 4.0 percent were on ramps and auxiliary lanes. Disabled vehicles of various types totaled 19 percent. Nondisabled vehicles totaled 81 percent. It usually was not known why the nondisabled vehicles stopped.

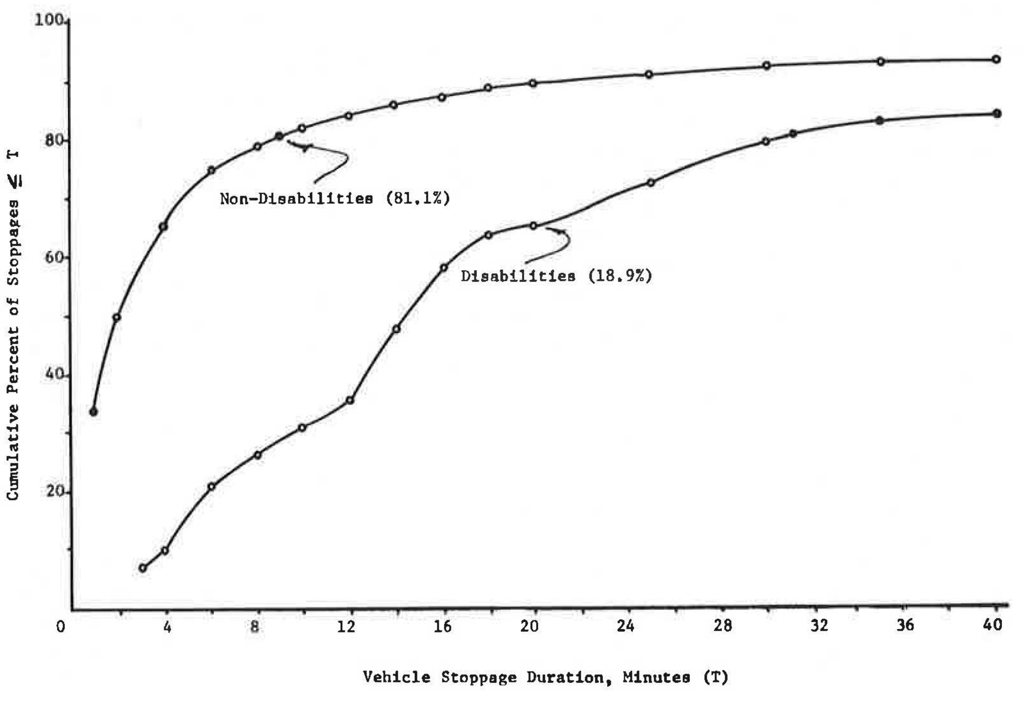
All of the nondisabled vehicles and 60 percent of the disabled vehicles required no assistance. Figure 6 shows the time duration of all stoppages; it suggests that any algorithm that initiates incident-response forces for noncritical short stoppages would introduce a high rate of operational false alarms into a system whose detection component has been optimized at the zero-false-alarm level.

The concept of detecting incidents through detecting individual stopped vehicles is based partially on safety considerations. Some idea of the magnitude of the hazard can be obtained from accident records. The 1972 Chicago-area accident records for 135 expressway miles (216 expressway kilometers) showed that only 52 out of 16,302 (0.3 percent) reported accidents involved vehicles on either the right or left shoulders. Of these 52, none involved fatalities, and only 9 occurred in the 1:00 to 6:00 a.m., low-volume period.

The costs of the proposed detection system should consider detection, verification, staffing, and response costs. Whenever an incident is detected, a response mechanism must begin. If closed-circuit television or other measures are to be used to help verify the nature of the incident, considerable additional costs are introduced. The proposed algorithms require detector and equipment factors representing, at 1,000-ft (304.8-m) spacings, about 15 times as many detectors as traffic stream monitoring in 1 lane at 0.5-mile (0.8-km) directional intervals.

All in all, the operational value of the proposed system appears to limit the application to those roadways, such as tunnels, where all stoppages block travel lanes (11).

Figure 6. Vehicle stoppage duration, cumulative frequency distribution.



REFERENCES

6. S. E. Bergsman and C. L. Shufflebarger, Jr. Shoulder Usage on an Urban Freeway. John C. Lodge Freeway Traffic Surveillance and Control Research Project, Study 417, Jan. 1962.
7. Disabled and Non-Disabled Vehicle Stoppage on the Congress Street Expressway. Illinois Department of Transportation, Chicago Area Expressway Surveillance Project Rept. 8, Nov. 1962.
8. A. Kuprijanow, S. Rosenzweig, and M. A. Warskow. Motorists' Needs and Services on Interstate Highways. NCHRP Rept. 64, 1969.
9. M. P. Pekala and J. M. McDermott. Summary Analysis: "Before" Study Data Collection for the I-80 Motorist Aid Communication System. Illinois Department of Transportation, Interim Rept., March 1974.
10. M. E. Goolsby and W. R. McCasland. Evaluation of an Emergency Call Box System. Texas Transportation Institute, Texas A&M Univ., Research Rept. 132-1F, Dec. 1969.
11. R. S. Foote. Intensive Traffic Management for Critical Roadways. Traffic Engineering and Control, Oct. and Nov. 1974.
12. F. DeRose, Jr. An Analysis of Random Freeway Traffic Accidents and Vehicle Disabilities. Highway Research Record 59, 1964, pp. 53-65.
13. M. E. Goolsby. Influence of Incidents on Freeway Quality of Service. Highway Research Record 349, 1971, pp. 41-46.

Harold J. Payne, Technology Service Corp.

Incident-detection algorithms generally fall into 2 classes: (a) those that are dependent on the propagation of a queue upstream of the incident site and (b) those that depend on a disruption in the pattern of traffic at successive detector stations. Most algorithms are members of the first class and find their greatest success under medium- and high-flow conditions, or, more precisely, under conditions in which the incident reduces the capacity to a level that is less than the approaching volume.

There are only a few representatives of the second class, and the paper of Dudek et al. is in this class. The earliest suggestion of an algorithm based on this principle is apparently that of Barker (14), and further work along these lines recently has been done by Sakasita and May (15). This class of algorithms is attractive for 2 reasons. First, it offers hope of detecting incidents under conditions in which the incident does not reduce the capacity below the level of approaching volume and hence is well suited for incident detection under low-volume conditions. Second, the phenomenon that is the basis for this class of algorithms is generally manifested in the data much sooner after the incident occurs than is the queue backup, which is the basis for the vast majority of presently available incident-detection algorithms.

Despite the attractive nature of this second class of algorithms, the development of effective algorithms that do not require an unreasonable number of detectors, such as detector stations placed at 500-ft (152.4-m) intervals, has proved to be quite difficult. A complicating operational problem, noted by Barker (14), is the presence of an on-ramp or an off-ramp between the detector stations. A further complication, discussed at some length by Dudek et al., is vulnerability to detector errors. The results presented by Sakasita and May (15) and by Dudek et al. appear promising, but it should be borne in mind that these results are based on simulations that do not consider the major problems presented by the complications cited.

The structure of the event-scan version of the algorithm presented in this paper does not appear to allow for modifications that might accommodate these practical considerations. In particular, there is no adjustable parameter to provide a trade-off between detection performance and false-alarm rate. In its present form, the algorithm is certainly not ready for operational use. Although the authors apparently recognize this, they do not indicate how the shortcomings can be overcome except to demand a heretofore unachieved quality in detectors.

Consideration of algorithms that identify a discrepancy between actual downstream flow conditions and a forecast of downstream flow conditions based on upstream flow conditions certainly is warranted. Development along these lines must proceed, however, with respect for known operational problems.

REFERENCES

14. J. L. Barker. Determination of Discontinuities in Traffic Flow as a Factor in Freeway Operational Control. Traffic Engineering, Nov. 1961, pp. 11-17.
15. M. Sakasita and A. D. May. Development and Evaluation of Incident Detection Algorithms for Electronic Detector Systems on Freeways. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, Spec. Rept., Aug. 1974.

DEVELOPMENT AND EVALUATION OF INCIDENT-DETECTION ALGORITHMS FOR ELECTRONIC-DETECTOR SYSTEMS ON FREEWAYS

Masami Sakasita, De Leuw, Cather and Company; and
Adolf D. May, Institute of Transportation and Traffic Engineering,
University of California, Berkeley

This paper describes a study of the development and evaluation of incident-detection algorithms for electronic-detector systems on freeways. The study was in 3 parts. The first part reviewed existing detection algorithms and the development of 2 new detection algorithms. During the development of these 2 algorithms, a section of a freeway lane formed by 2 detectors at both ends was treated as a system, and an attempt was made to express traffic movements by using dynamic equations. The second part involved the development of a microscopic simulation model of freeway traffic performance. The simulation model was capable of simulating traffic conditions on a freeway under incident and nonincident situations. The output of the simulation model, which was recorded at presence detectors in each lane at 0.125-mile (0.20-km) spacings, was stored on a magnetic tape and played back later to test each detection algorithm. The third part evaluated the newly developed algorithms. The California model, which is considered to be the most widely known algorithm, was compared to the 2 proposed algorithms.

•INCIDENT-FREE flow conditions on urban freeways are abnormal. Previous studies have noted that freeway incident rates may be as high as 1 incident per directional mile (0.62 incident per directional kilometer) per hour (1). Freeway incidents are hazardous to all in the traffic stream. And reduction in freeway capacity may be significant enough to cause congestion and thus create further hazards and delay for passing motorists.

Knowledge of incidents that reduce capacity is extremely important for freeway-control strategies. Previous work has been undertaken by us and by others to evaluate incident-detection systems (2, 3, 4). But the development and evaluation of electronic-detector systems have received only limited attention. These systems generally are based on uncontrolled empirical experiments only. The results of these empirical experiments are discussed in the paper.

This paper describes a study of the development and evaluation of incident-detection algorithms for electronic-detector systems. The study consisted of the development of 2 detection algorithms and a simulation model of freeway traffic performance. In addition, the algorithms were evaluated by using the simulation model, TRAFFIC. Two detection algorithms were proposed. One was the dynamic model, which applies the information-theory technique (5). The other was the stream discontinuity model, which is based on a macroscopic treatment of the dynamic model. This study also evaluated the detection algorithms. The California model, which is considered to be the most widely known, was evaluated along with the 2 newly developed algorithms.

AUTOMATIC DETECTION ALGORITHMS

Several unique studies of automatic detection systems have been completed. This paper will describe 2 operational systems and 2 experimental systems.

Port of New York Authority Method

The Port of New York Authority method is described in detail elsewhere (7, 9). Control strategy for the Lincoln Tunnel is based on the number of vehicles in each of 3 sections. If y_k = the number of vehicles in a section at the beginning of the kth observation period, u_{1k} = the number of vehicles that have entered the section in the kth time period, and u_{2k} = the number of vehicles that have left the section in the kth time period, then the number of vehicles at the beginning of the k + 1 period should be

$$y_{k+1} = y_k + u_{1k} - u_{2k} \qquad (1)$$

The traffic density in each section of the tunnel was estimated by this relation, and incidents were predicted for abnormally large values of density. This method, however, accumulates errors because of miscounts by the sensors. Therefore, it requires microscopic identification of vehicles after a certain time. This method can be used in the tunnel because of the tunnel's accurate detection system and because vehicles are not permitted to change lanes in the tunnel. Recently this method was improved by applying the extended Kalman filtering theory to it.

The California Model

The California model (6), which has produced promising results, uses occupancy as a measure of traffic conditions. It is less accurate than that used in the Lincoln Tunnel in New York. The basic concept of the method is a comparison of occupancies at the neighboring detectors in the same time interval and a comparison of occupancies at adjacent time intervals at the same detector. Occupancies are 1-min values updated every 20 or 30 sec. The detection criterion should predict an incident when all the calculated values of X_1, X_2, and X_3 exceed K_1, K_2, and K_3 at the same time.

$$(OCC_i)_t - (OCC_{i+1})_t = X_1 \geq K_1 \qquad (2)$$

$$\frac{(OCC_i)_t - (OCC_{i+1})_t}{(OCC_i)_t} = X_2 \geq K_2 \qquad (3)$$

$$\frac{(OCC_{i+1})_{t-B} - (OCC_{i+1})_t}{(OCC_{i+1})_{t-B}} = X_3 \geq K_3 \qquad (4)$$

where

OCC_i = occupancy at station i that is counted in the direction of travel,
t = time instant,
X_1, X_2, X_3 = X occupancy values,
K_1, K_2, K_3 = K occupancy values (K_1, K_2, and K_3, which can be adjusted depending on location, are 8, 0.55, and 0.10 respectively for the nonpeak period and 8, 0.55, and 0.15 respectively for the peak period), and
B = a time period of 20 or 30 sec (backward shift operator).

Texas Transportation Institute Method

Experimental studies conducted by the Texas Transportation Institute (8) suggest several new approaches to automatic detection systems, one of which is the kinetic energy approach. Kinetic energy is computed as follows.

$$E \propto k\mu^2$$

that is,

$$E \propto q^2/\theta$$

where

E = kinetic energy,
k = density,
μ = speed,
q = flow, and
θ = occupancy.

In a study based on this approach, the 1-min kinetic energy values were compared with preestablished limits and a probable incident was reported whenever the measurements exceeded their lower limits. This approach was extended to an individual-lane-energy approach that was done lane by lane. The results of the experimental study tended to show that this method has a high false-alarm rate.

Double Exponential Smoothing Method

Cook and Cleveland (10) introduced another new approach for automatic incident-detection: a time-series analysis technique called the double exponential smoothing method. In this method, the smoothing function of observation at the tth time increment, $S_t(x)$, is expressed as

$$S_t(x) = \alpha x_t + \beta S_{t-1}(x) \tag{5}$$

where

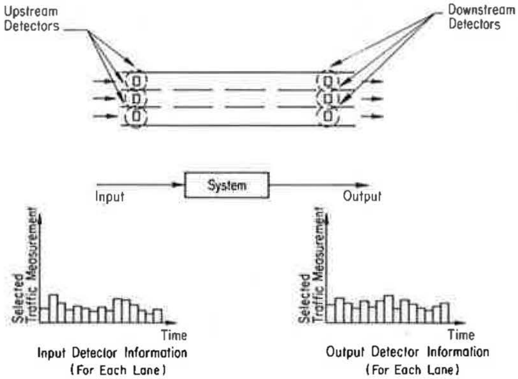
α = smoothing constant,
x_t = tth observation, and
β = 1 - α.

Traffic parameters were forecast, and incidents were predicted for parameter values that exceeded the preset limit based on Eq. 5. Experiments have shown that this method is comparable or superior to the California model.

DEVELOPMENT OF INCIDENT-DETECTION ALGORITHMS

To develop incident-detection algorithms, one should treat a section of freeway lane formed by 2 detectors at both ends as a system. An upstream detector would provide the system with certain output traffic measures. This system is shown in Figure 1. Inside the system, vehicles do or do not change lanes. The function of the system is expressed by a travel-time distribution.

Figure 1. Freeway system.



Automatic Incident-Detection Concept

Traffic phenomena after an incident occurs differ, depending on whether the traffic-flow rate is larger than, smaller than, or equal to the reduced capacity. When only local disturbances are observed, shockwaves propagate neither upstream nor downstream of the incident. When the input flow rate is larger than reduced capacity, shockwaves propagate both upstream and downstream of the incident; consequently, extreme changes occur in traffic-flow measurements, and these changes permit one to look at either individual detectors or a pair of neighboring detectors on the freeway. When significant changes are observed in the detector measurements, an incident is predicted. On the other hand, when the input flow rate is smaller than capacity, sudden changes in the system do not occur. This phenomenon requires the automatic-detection algorithm to detect local disturbances that do not have the characteristics of extreme changes.

Most detection algorithms, including the California model, predict incidents by observing sudden changes in detector measurements. In this paper, an effort has been made to develop detection algorithms that detect accurately not only the occurrence of incidents at situations when flow is higher than reduced capacity but also incidents in the lower flow level, particularly when this level is far less than reduced capacity. Two detection algorithms are proposed in this paper. The first, the dynamic model, deals with the impulse-response function of the system, which is interpreted as a probability density function of travel time between 2 detectors; the second, the stream discontinuity model, tries to detect discontinuities in a traffic stream by comparing occupancies at 2 locations on a freeway by using proper time shifts in measuring occupancies.

Dynamic Model

It is assumed that the arrival pattern of vehicles on a freeway lane remains essentially unchanged under nonincident conditions. It is also assumed that the occurrence of an incident creates disturbances in the arrival pattern downstream of the incident location. In developing the dynamic model the first task is to formulate a model that can represent the traffic phenomena previously described.

Formulation of the Model

Consider a longitudinal pair of detectors in a freeway lane. Traffic information (such as occupancy or number of vehicles) is obtained at each detector for each equally spaced time interval. Using the sampling intervals as the unit of time, one denotes input information and output information at t as X_t and Y_t respectively. If all the vehicles travel with the same speed and no lane changing takes place between the 2 detectors, then the input and output have a relationship such that

$$Y_t = X_{t-\tau} \tag{6}$$

where τ = travel time (time shift) between the 2 detectors. But, in the real world, all vehicles do not travel at the same speed or without changing lanes. So the relationship between the input and output is a little more complicated. Let us define X_t and Y_t as the respective deviations of t from the mean of input and output information over a long time period. Then output of the system Y_t is represented as a linear aggregate of input deviations at t, t - 1, t - 2, . . . , and the noise sequence, N_t, is such that

$$\begin{aligned} Y_t &= v_o X_t + v_1 X_{t-1} + v_2 X_{t-2} + \ldots + N_t \\ &= (v_o + v_1 B + v_2 B^2 + \ldots) X_t + N_t \\ &= v(B) X_t + N_t \end{aligned} \qquad (7)$$

In the field of time-series analysis, the weights of v_o, v_1, . . . in Eq. 7 are the impulse-response function of the system and v(B) is the transfer function of the filter. B is defined by $BX_t = X_{t-1}$. The sum of the weights v_o, v_1, . . . is equal to the steady-state gain of the system, g.

$$\sum_{i=0}^{\infty} v_i = g \qquad (8)$$

The steady-state gain inthis application is interpreted as the ratio of the mean values of traffic information at the upstream and downstream detectors. Under nonincident conditions g is assumed to come very close to 1, and v_i represents the travel-time probability between the 2 i detectors.

N_t in Eq. 7 corrupts the linear dynamic system; this is mainly due to lane changing and detector errors. It is assumed that X_t is uncorrelated with N_t.

If it is known that the v_is are effectively zero beyond i = K, then, in order to determine the v_is of the system $(N + 1) \geq (K + 1)$, sets of X_t and Y_t have to be made. Observations would yield equations of the form

$$\begin{gathered} x_t v_o + x_{t-1} + \ldots + x_{t-K} v_K = y_t \\ x_{t-1} v_o + x_{t-2} v_1 + \ldots + x_{t-K-1} v_K = y_{t-1} \\ \ldots \\ x_{t-N} v_o + x_{t-N-1} + \ldots + x_{t-K-N} v_K = y_{t-N} \end{gathered} \qquad (9)$$

If only K + 1 sets of measurements are made, then unique solutions would exist for v_i, but noise and measurements errors would cause Eq. 9 to yield incorrect values. Thus more measurements than the number of unknowns are taken. If substitution of the values $\hat{v}_o, \hat{v}, \ldots, \hat{v}_K$ for the unknown $v_o, v, \ldots, v_K$ on the left side of Eq. 9 yields $\hat{y}_t, \hat{y}_{t-1}, \ldots, \hat{y}_{t-N}$, which differ from $y_t, y_{t-1}, \ldots, y_{t-N}$ by $e_i = \hat{y}_i - y_i$ (where i = t, . . . , t - N), then $\hat{v}_i$ is to be determined such that the $\hat{v}_i$s have the smallest mean square deviation; that is to say that

$$\sum_{i=t-N}^{t} e_i^2 = \sum_{i=t-N}^{t} (\hat{y}_i - y_i)^2 \qquad (10)$$

is a minimum.

If the vectors $\hat{V}$ and $\underline{Y}$ and the matrix X are defined such that

$$\underline{\hat{V}} = \begin{bmatrix} \hat{v}_0 \\ \hat{v}_1 \\ \cdot \\ \cdot \\ \cdot \\ \hat{v}_K \end{bmatrix} \tag{11}$$

$$X = \begin{bmatrix} x_t & x_{t-1} & \cdots & x_{t-K} \\ \cdot \\ \cdot \\ \cdot \\ x_{t-N} & x_{t-N-1} & \cdots & x_{t-N-K} \end{bmatrix} \tag{12}$$

$$\underline{Y} = \begin{bmatrix} y_t \\ y_{t-1} \\ \cdot \\ \cdot \\ \cdot \\ y_{t-N-1} \end{bmatrix} \tag{13}$$

then the $\underline{\hat{V}}$ that gives the smallest possible mean square deviation is given as

$$\underline{\hat{V}} = (X^T X)^{-1} X^T \underline{Y} \tag{14}$$

where T = average travel time of vehicles in each T-sec interval. For a large value of N, Eq. 14 can be rewritten

$$\underline{\hat{V}} = C_{xx}^{-1} C_{xy} \tag{15}$$

where

$$C_{xx} = \begin{bmatrix} c_{xx}(0)c_{xx}(1)\ldots c_{xx}(K) \\ c_{xx}(1)c_{xx}(0)\ldots c_{xx}(K-1) \\ \cdot \\ \cdot \\ \cdot \\ c_{xx}(K)c_{xx}(K-1)\ldots c_{xx}(0) \end{bmatrix} \tag{16}$$

$$C_{xy} = \begin{bmatrix} c_{xy}(0) \\ c_{xy}(1) \\ \cdot \\ \cdot \\ \cdot \\ c_{xy}(K) \end{bmatrix} \tag{17}$$

$c_{xx}(j)$ and $c_{xy}(j)$ are the respective estimates of the autocovariance coefficient of the x series and the cross-covariance coefficient between x and y that are given as

$$c_{xx}(j) = \frac{1}{N} \sum_{i=t-N}^{t} x_i x_{i-j} \tag{18}$$

$$c_{xy}(j) = \frac{1}{N} \sum_{i=t-N}^{t} x_{i-j} y_i \tag{19}$$

If the input series is not autocorrelated, then matrix C_{xx} can be considered as a diagonal matrix, and $\underline{V}$ would be given as

$$\hat{\underline{V}} = \frac{1}{\sigma_2^2} C_{xy} \tag{20}$$

$$\hat{v}_i = \frac{1}{\sigma_2^2} c_{xy}(i) = \frac{\sigma_y}{\sigma_x} r_{xy}(i) \tag{21}$$

where $r_{xy}(i)$ = estimate of the cross-correlation coefficient at lag i.

If the input series is autocorrelated, then either Eq. 14 can be computed directly or the input series can be prewhitened in obtaining the $\hat{v}_i$s. But prewhitening the input series considerably simplifies the solution process. As shown in the next section, the autocorrelation functions of several input series that are obtained from TRAFFIC were computed, and it was found that these input series were not autocorrelated. The dynamic model does not have a prewhitening routine, and all input series are treated as nonautocorrelated series.

Detection Criterion

It is assumed that each vehicle's travel time between the 2 detectors under a certain traffic condition follows a certain distribution. Observed sets of v_is are considered similar to each other. If the observations of v_is are performed for only certain time periods that would effectively cover the necessary range of the impulse-response function under the nonincident condition, then the sum of v_is would be close to 1, but the sum of v_is under incident conditions would become much smaller than 1 because travel-time distribution would be disturbed by the existence of an incident. At the same time a large difference is assumed to be observed in values of upstream and downstream measurements if the proper shifted τ is used in measuring them. Let

$$\alpha = \sum_{V \text{ observed } i} \hat{v}_i \tag{22}$$

$$\beta = \sum_{i=t+\tau}^{t+N+\tau} x_i \Bigg/ \sum_{i=t}^{\tau+N} y_i \tag{23}$$

where N + 1 = number of input and output noise observations. The detection criterion should predict an incident for a small value of α and a large value of β. Or, more precisely, an incident should be predicted if an observed α, β point is in the critical region that is shown in Figure 2. The α_0 and β_0 values in Figure 2 were determined empirically.

Numerical Example

Before applying the dynamic model to a simulated traffic condition, we first observed whether the input series was autocorrelated. Estimated autocorrelation functions in the right lane at input levels of 400, 1,000 and 1,600 vehicles per hour (vph) per lane were obtained. The study showed that the input series was not autocorrelated in most cases.

Figure 3 shows the estimated cross-correlation function of the vehicle arrival counts at 2 detectors located in the right lane that are 0.125 mile (0.20 km) apart. Because observed variances of the input and output series were nearly equal in all the 3 flow levels, the estimated cross-correlation function was almost identical to the impulse response function of the system. There were 80 upstream and 80 downstream 1-sec arrival counts. No incident occurred when the cross-correlation function was observed. The flow level was 1,000 vph per lane. τ (estimated from the downstream detector) was 8.1 sec. Here, v_7, v_8, v_9, and v_{10} formed the observed travel-time distribution.

Figure 4 shows the estimated cross-correlation function of the arrival counts at the same 2 detectors but with an incident located between them. Again there were 80 upstream and 80 downstream 1-sec arrival counts, and the flow level was 1,000 vph per lane. Figure 4 shows that no $\hat{v}_t$ values, when compared to standard error, are significantly large.

Figure 5 shows the plot of α, β values in the incident lane under both nonincident and incident conditions. Flow was 1,000 vph per lane, and detectors were spaced 0.125 mile (0.20 km) apart. It is evident that the points under the incident condition are distinctly separate from the points under the nonincident condition as shown by the ellipse in the figure.

Stream Discontinuity Model

The dynamic model estimates by the least squares method the probability function of the travel time between the longitudinally placed detectors. It is assumed that, if an incident occurs, the probability function of travel time would be disturbed. In the stream discontinuity model, average travel time instead of travel-time distribution is estimated, and it is assumed that the occurrence of an incident creates a large difference in traffic measurements at the upstream and downstream detectors.

Formulation of the Model

Consider a pair of detectors in a lane on a freeway; occupancies are measured at each detector for equally spaced time intervals. Occupancies at 2 detectors in a T ending at t are expressed by $\theta_{x,t}$ and $\theta_{y,t}$ where

$\theta_{x,t}$ = upstream detector occupancy in seconds, and
$\theta_{y,t}$ = downstream detector occupancy in seconds.

If the ending time of each T at the upstream detector is shifted by τ, then $\theta_{x,t}$ and $\theta_{y,t}$ are assumed to differ greatly under incident condition. In a similar manner to that of the California model, the stream discontinuity model considers 2 measures such that

$$Z_1 = \theta_{y,t} - \theta_{x,t-\tau} \tag{24}$$

and

$$Z_2 = \theta_{y,t}/\theta_{x,t-\tau} \tag{25}$$

Figure 2. Critical region of the dynamic model.

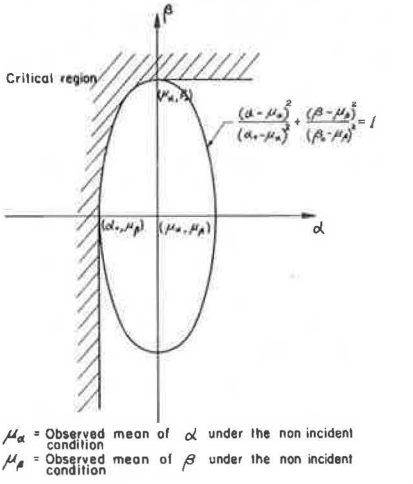


Figure 3. Estimated cross-correlation function in right lane under normal condition.

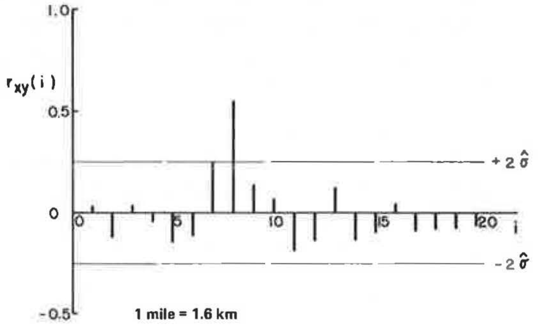


Figure 4. Estimated cross-correlation function in right lane under incident condition.

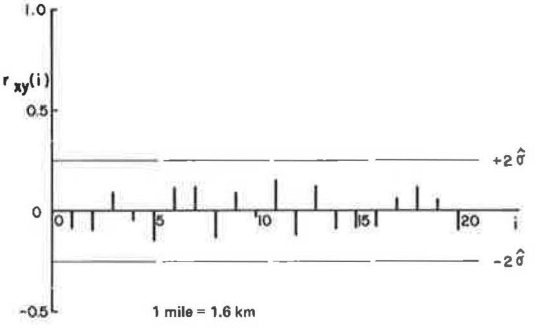
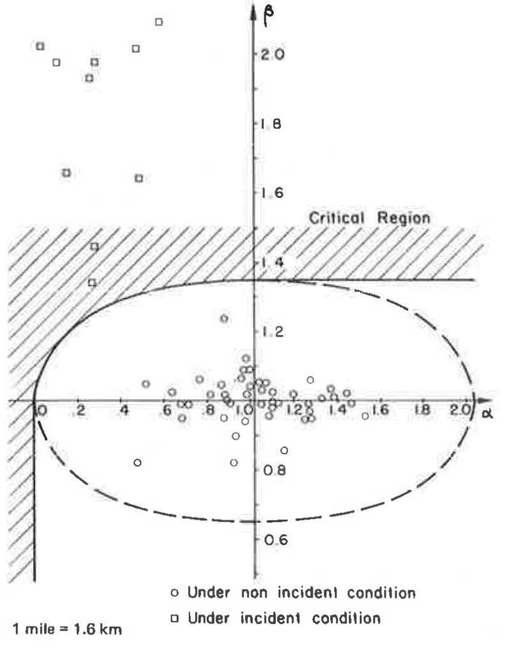


Figure 5. α, β values.



The detection criterion should predict an incident when calculated values of Z_1 and Z_2 exceed Z_1^* and Z_2^* at the same time. The values Z_1^* and Z_2^* are determined empirically.

T can vary depending on detector spacing. Larger T values would give stable results but larger detection times. On the other hand, smaller T values would give shorter detection times but more false alarms.

The amount of shifted τ is estimated from downstream detector information. τ is given as

$$\tau = \frac{\theta_{\gamma,t}}{f} \times \frac{L_d}{\ell'} \tag{26}$$

where

L_d = detector spacing in feet (meters),
ℓ' = average vehicle length plus effective detector length in feet (meters), and
f = number of vehicles counted at downstream detector in T.

Numerical Example

This example was taken from a simulation result produced by TRAFFIC. The flow level was 1,000 vph per lane. An incident was generated in the right lane of a 3-lane freeway 10 min after the simulation began. The detectors were set 0.25 mile (0.40 km) apart. T was 60 sec.

Figure 6 shows the plot of observed τ, Z_1, and Z_2. It is clearly seen that both Z_1 and Z_2 values were stable before the incident occurred, but they became unstable after the incident. Because shifted τ, which represents estimated average travel time between the 2 detectors along the freeway lane, is estimated from the downstream detector, the value of τ tends to be smaller after the incident.

DEVELOPMENT OF A SIMULATION MODEL

TRAFFIC Simulation Model

A simulation model, TRAFFIC, was developed to evaluate various incident-detection schemes. TRAFFIC is a microscopic Monte Carlo simulation model. The program consists of 12 subprograms and 11 functions. Its program length is about 1,800 statements, and it uses about 36,000 octal core locations. Simulation was carried out for a 1.5-mile (2.4-km) section, but the first 0.5-mile (0.8-km) section was used for warm-up, so no output was obtained from this subsection. The physical structure of the freeway section and detector arrangements are shown in Figure 7. The detectors were uniformly spaced 0.125 mile (0.20 km) apart, and each transverse lane had a separate detector. Traffic information was obtained through these detectors and stored on a magnetic tape. The information from each detector consisted of occupancies (or pulse lengths) and actual speeds of vehicles. This information was the input for the detection algorithm programs. By storing the traffic information on tape, one is able to avoid repetitious runs of the simulation program. The scanning time of the model is 1 sec, which is considered to be an allowable maximum value for this type of freeway simulation.

Performance Analysis

To check the reasonableness of the simulated traffic performance, we analyzed the output data from 9 selected detectors and compared the results to field measurements.

Analyzed items included spot-speed distribution, headway distribution, volume-density relationship, lane-changing phenomena, queue evolution, and capacity estimation.

Analysis of the output shows that the simulation model gave realistic results. The simulation results are shown in Figures 8, 9, and 10. Figure 8 shows the speed distribution at 3 flow levels; speed distribution curves from the Highway Capacity Manual (11) are superimposed on the figure's 3 graphs. At the 400 vph per lane flow level the speed distribution of the simulation was almost the same as the Highway Capacity Manual's distribution. But at the 1,000 vph per lane and 1,600 vph per lane flow levels, simulation results differed greatly from the distributions of the Highway Capacity Manual. But it should be noted that on newly constructed freeways the mean of the speed distributions at these flow levels is even higher than simulation results. Makigami, Woodie, and May (12) noted that on the East Bayshore Freeway in the San Francisco Bay area the observed mean speed for 400 vph per lane was 95.3 fps (29.07 m/s); for 1,000 vph per lane it was 86.5 fps (26.38 m/s) and for 1,600 vph per lane it was 83.6 fps (25.50 m/s). Figure 9 shows headway distributions for these 3 flow levels. These distributions are compared in Figure 9 to the observed distributions noted by May and Wagner (13). Simulation results reasonably fit the observed distributions. Figure 10 shows the volume-density diagram obtained from the 3 simulation runs. The resulting curve, which was drawn from 12 observations, appears reasonable.

Transition matrices that show lane-changing phenomena under nonincident condition were constructed and compared to the field observation values reported by Worrall, Bullen, and Gur (15); it was found that the simulation model gave reasonable results. Lane-changing phenomena under incident conditions were not tested because empirical data were not available. Queue evolution was observed at the 1,600 vph per lane flow level; it was found that the average speed of the queue front was 103 ft/min (0.523 m/s) or 1.2 mph (1.9 km/h). The capacity was found to range from 2,200 vph per lane to 2.300 vph per lane as shown in Figure 10. The reduced capacity (calculated for the 1,600 vph per lane level) was 4,518 vph or 2,259 vph per lane. The capacity values showed reasonable results.

EVALUATION OF THE DETECTION ALGORITHMS

Production runs were made with the simulation model for the 3 traffic levels (400, 1,000, and 1,600 vph per lane) and for 2 different incident occurrences (an incident in the right lane and an incident in the middle lane). Each simulation run was conducted in 20 min of real time. Approximately 10 min after the beginning of the simulation run, an incident was generated in the right lane (or middle lane) of the freeway midway between the 5th and 6th detector sets. This simulation procedure provided a simulation run of 10 min before the incident and 10 min after the incident. Detector information from all the detectors was stored on a magnetic tape.

The aforementioned detection algorithms were computerized and evaluated by using the traffic performance on the simulation runs. The California model was computerized and compared to the newly developed models.

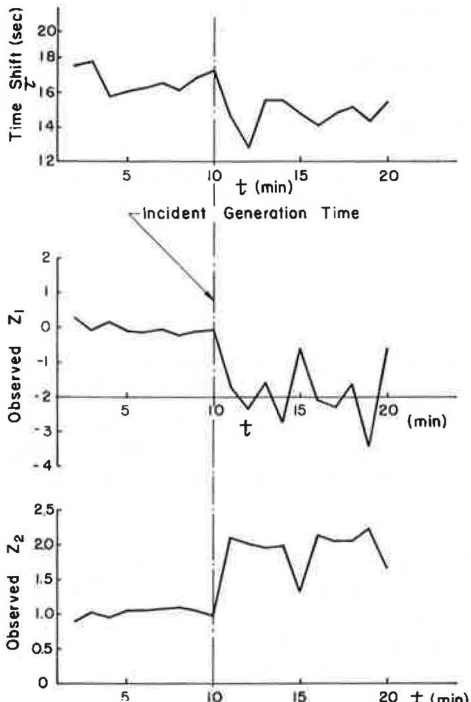
Experiment Design

The variables considered in designing the experiment were as follows:

1. Detection algorithms (California, dynamic, and stream discontinuity models);
2. Detection configurations [0.125-mile (0.20-km), 0.25-mile (0.40-km), 0.5-mile (0.80-km), and 1-mile (1.6-km) spacings (Fig. 11)];
3. Traffic-flow levels (400, 1,000, and 1,600 vph per lane); and
4. Incident location [2 locations: right lane and middle lane, both of which were 2,970 ft (990 m) from origin of effective simulation section].

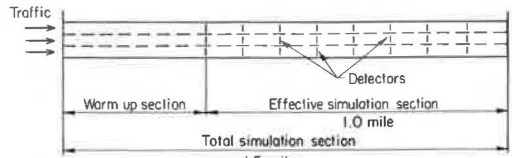
Combinations of these variables made 36 experiments for each algorithm. Evaluation of algorithms and detector spacings was based on these experiments.

Figure 6. Observed τ, Z_1, Z_2 values from the simulation results.



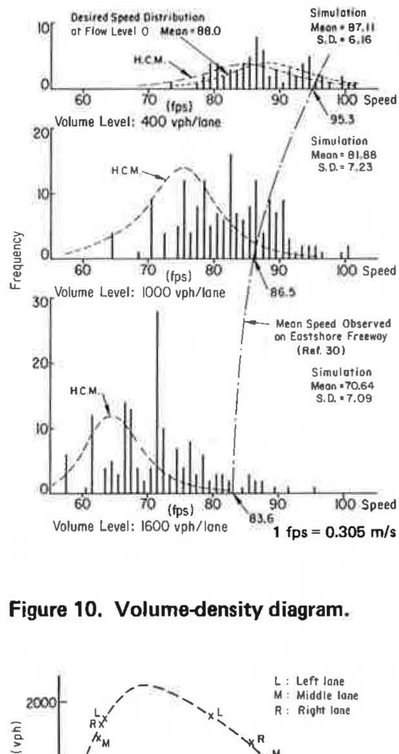
1 mile = 1.6 km

Figure 7. Detector arrangement on the freeway section.



1 mile = 1.6 km

Figure 8. Spot-speed distributions at the 3 flow levels, middle point of middle lane.



1 fps = 0.305 m/s

Figure 9. Headway distributions at the 3 flow levels, middle point of middle lane.

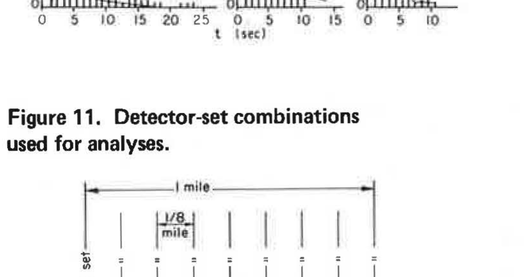
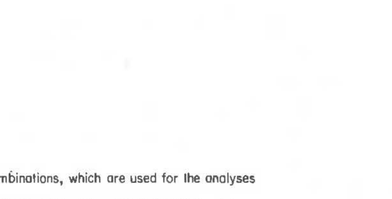
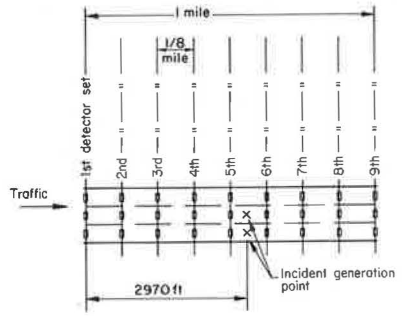


Figure 10. Volume-density diagram.



1 vehicle/mile = 0.63 vehicle/km

Figure 11. Detector-set combinations used for analyses.



Evaluation Criteria

The criteria of evaluation considered in this experiment were as follows:

1. Probability of no detection, p_{nd}, which is the probability of having no alarm indicating an incident in the 10 min after an incident;
2. Average detection time, t_d, which is calculated; and
3. Average number of detections in 10 detection trials (n_d). In each incident case, the detection of the incident was tried 10 times after the incident generation to show how many times an incident is detected in the 10 trials. A large n_d value would indicate that the algorithm is highly reliable.

False-Alarm Probability

False-alarm probability (p_{fa}) is the probability of having a false alarm at a detection trial when no incident is on the freeway section. False-alarm probability is related directly to the critical values of each detection algorithm. This probability can almost be controlled by changing the critical values. Although p_{fa} is often one of the evaluation criteria of incident-detection algorithms, it is treated not as an evaluation criterion but as a controllable variable in this experiment. The algorithms are evaluated for the p_{fa} of 0.001.

In reality, this p_{fa} should be different depending on the number of detectors used in the surveillance system and the number of detection trials per unit of time. For example, if the number of detectors used is 100, and detection trials are performed each minute, then 6,000 detection trials would be performed in 1 hour. If a p_{fa} of 0.001 is used, then the expected number of false alarms in 1 hour in the system would be 6.

Detection Algorithm Results

The experiment was initiated by using several critical values set up for each incident-detection algorithm. The resulting operating characteristics of the 3 algorithms are shown in Figure 12. In Figure 12, the vertical axis represents p_{nd}, and the horizontal axis represents p_{fa}. In any incident-detection algorithm, there is a trade-off between p_{nd} and p_{fa}. For example, at the 0.00 p_{fa} level, the California model has a probability of 0.40 of not detecting an incident; the stream discontinuity model has a lower value of 0.19; and the dynamic model has the lowest value of 0.07. To achieve the 0.00 probability of no detection, the California model has to allow the highest false-alarm probability of 0.031; to achieve the same level, the dynamic model only has to allow 0.012. Obviously the dynamic model shows the best result.

Comparison of Detection Algorithm Results

At the false-alarm probability 0.001, a comparison of detection algorithms was made for the t_d for each flow level and n_d. Figure 13 shows the comparison at the 3 flow levels. t_d was calculated for each flow level and detector spacing. In calculating t_d, no detection was counted as 11 min of detection time. In Figure 13, the observed points that contain no detections are shown.

At the flow level of 400 vph per lane, the effect of detector spacing on detection time was not strong. Because of no-detection observations in the original data, a straightforward comparison is difficult. But it can be seen that at the 0.125-, 0.25-, and 0.5-mile (0.20-, 0.40-, and 0.80-km) detector spacings the dynamic model had the best results; at the 1-mile (1.6-km) detector spacing the dynamic model again showed the best results. The stochastic elements of the traffic flow prevented any monotonic trend in the curves for any of the 3 algorithms.

At the 1,000 vph per lane flow level, a monotonic increase of the average detection time was observed in all of the 3 algorithm results. In this case also, the dynamic model had the best results for t_d except at the 0.25-mile (0.40-km) detector spacing.

At the 1,600 vph per lane flow level, the t_ds of the 3 algorithms increased monotonically as a function of the detector spacing. The dynamic model showed the best result at all 4 detector spacings. It is rather surprising that the California model showed very poor results at the 0.5- and 1- mile (0.40- and 0.80-km) detector spacings.

Looking at the 3 graphs in Figure 13, one should notice that especially the California model shows rather unpredictable results. This may indicate that the California model tends to pick up stochastic elements of the traffic flow more easily than the other 2 algorithms.

Figure 14 shows the n_d trials under the same false-alarm level (p_{fa} = 0.001). At the 400 vph per lane flow level, the dynamic model showed the best results.

At the 1,000 vph per lane and 1,600 vph per lane flow levels, the dynamic model and the stream discontinuity model showed better results compared to the California model. The number of detections in 10 min tended to decrease as space increased except for the flow level of 400 vph per lane for the California model.

CONCLUSIONS

Two detection algorithms were proposed and tested with the microscopic simulation model that was developed to analyze detector schemes. The California model was compared to these 2 detection algorithms, and they compared favorably at all flow levels, particularly when detectors were spaced far apart.

The results of this study have revealed the influence of detector spacings and flow levels on the 3 detection algorithms. Further research is required to obtain more comprehensive results and to perform more exhaustive evaluations of these and other possible detection algorithms. The simulation model is limited in terms of its geometrics, demand patterns, and its ability to change capacity and demand over time and over the length of route. However, these limitations are not inherent in the methodology. More flexible models can be constructed, and other detection algorithms can be developed based on the results given in this paper. In addition, some modifications in the methodology such as reducing the decision interval for incident prediction and additional and longer simulation runs are desirable. Finally, field experiments should be conducted to validate the results of this study under real-life situations.

ACKNOWLEDGMENT

We wish to thank William A. Stock for his assistance in computer programming.

REFERENCES

1. S. J. Kolenko and J. A. Albergo. Disabled and Non-Disabled Vehicle Stoppage on the Congress Street Expressway. Chicago Area Expressway Surveillance Project, Illinois Department of Transportation, 1962.
2. H. Keller. An Analysis of the Freeway Emergency Service Systems. Operations Research Center, Univ. of California, Berkeley, ORC 69-20, Aug. 1969.
3. M. Sakasita, C. K. Lu, and A. D. May. Evaluation of Freeway Emergency Service Systems Using a Simulation Model. Operations Research Center, Univ. of California, Berkeley, ORC 70-23, July 1970.
4. W-M. Chow and A. D. May. Searching for the Best Locations of Service Facilities Along the Freeway. Operations Research Center, Univ. of California, Berkeley, ORC 72-19, Aug. 1972.
5. G. E. Box and G. Jenkins. Time Series Analysis Forecasting and Control. Holden Day, 1969.
6. R. Kahn. Interim Report on Incident Detection Logics for the Los Angeles Freeway

Figure 12. Operating characteristics of the 3 algorithms.

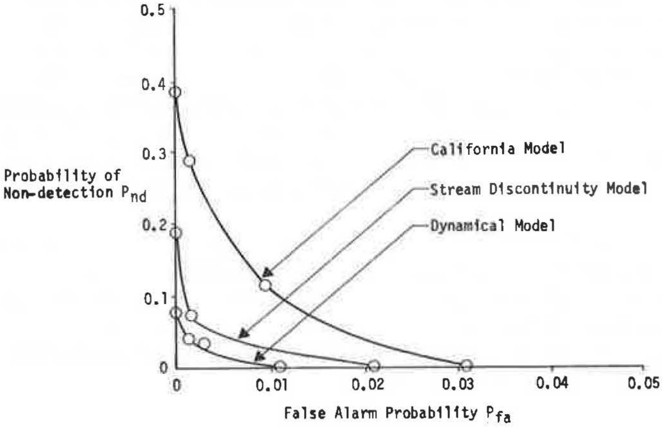


Figure 13. Comparison of average detection times, false-alarm probability: 0.001.

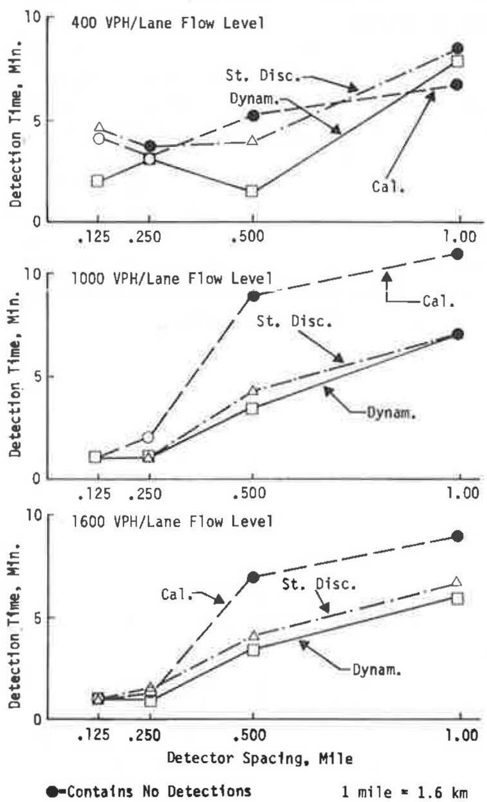
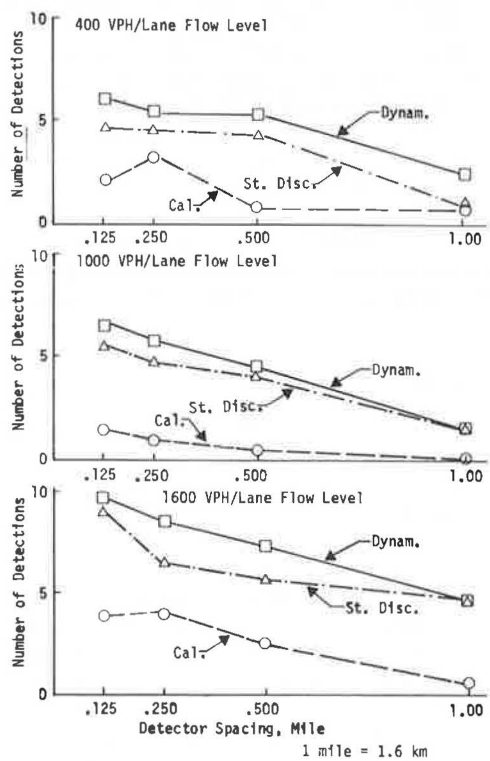


Figure 14. Comparison of average number of detections in 10 detection trials, false-alarm probability: 0.001.



Surveillance and Control Project. California Division of Highways, Sept. 1972.
7. D. C. Gazis and R. S. Foote. Surveillance and Control of Tunnel Traffic by On-Line Digital Computer. IBM Research, Jan. 1969.
8. K. G. Courage and M. Levin. A Freeway Corridor Surveillance, Information, and Control System. Texas Transportation Institute, Texas A&M Univ., Research Rept. 488-8.
9. M. W. Szeto and D. C. Gazis. Application of Kalman Filtering to the Surveillance and Control of Traffic Systems. Transportation Science, Vol. 6, 1972, pp. 419-439.
10. A. R. Cook and D. E. Cleveland. Detection of Freeway Capacity-Reducing Incidents by Traffic-Stream Measurements. Transportation Research Record 495, 1974, pp. 1-11.
11. Highway Capacity Manual. HRB Spec. Rept. 87, 1965.
12. Y. Makigami, L. Woodie, and A. D. May. Analytic Techniques for Evaluating Freeway Improvements. In Bay Area Freeway Operations Study, Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, Vol. 3, Aug. 1970.
13. A. D. May and F. A. Wagner. Headway Characteristics and Interrelationships of Fundamental Characteristics of Traffic Flow. HRB Proc., Vol. 39, 1960.
14. J. L. Barker. Determination of Discontinuities in Traffic Flow as a Factor in Freeway Operational Control. Traffic Engineering, Nov. 1961.
15. R. D. Worrall, A. G. R. Bullen, and Y. Gur. An Elementary Stochastic Model of Lane-Changing on a Multilane Highway. Highway Research Record 308, 1970, pp. 1-12.
16. M. Sakasita. Development and Evaluation of Incident Detection Algorithms for Electronic Detector Systems on Freeways. Univ. of California, Berkeley, PhD dissertation, July 1974.

EVALUATION OF A PROTOTYPE WARNING SYSTEM FOR URBAN FREEWAYS

Conrad L. Dudek, R. Dale Huchingson, and Gene Ritch,
Texas Transportation Institute, Texas A&M University

This paper evaluates a prototype real-time system that warns approaching motorists of stoppages downstream of cresting vertical curves. Before and after studies were conducted to measure primary and secondary accidents. A questionnaire survey was administered to obtain motorist reactions. The study results revealed that the warning system is cost effective. Both primary and secondary accidents were reduced. The results of the questionnaire study indicated that motorists believed that the system was useful, that the warning sign was readily noticed, and that the message was generally understandable and appropriate to traffic conditions.

•AN EXPERIMENTAL warning system has been in operation on the inbound control section of the Gulf Freeway in Houston since April 3, 1972 (1, 2). The purpose of the system is to assist freeway drivers approaching cresting vertical curves by giving them information on the downstream traffic flow and by alerting them to stoppage waves downstream of the crest.

Three overpasses were selected as the sites for pilot installations to study the effectiveness of the warning system, develop automatic-control algorithms, and further evaluate the design concepts. The system consisted of a static sign with attached flashing beacons (Fig. 1) located upstream of each overpass crest and a flashing beacon mounted on the bridge rail on the top of each crest (Fig. 2). The warning signs were controlled automatically by a digital computer. Double-loop detectors were installed on each lane and located on both sides of the 3 overpasses. The primary function of the detectors downstream of the overpass was to sense stoppage waves to activate the warning sign. The upstream detectors would indicate the time that the sign should be turned off. Installation sites and the freeway sections influenced by the 3 warning signs are shown in Figure 3.

This paper evaluates the prototype warning system.

METHOD OF STUDY

Measures of Effectiveness

The objective of the warning system is to alert approaching drivers of stoppages downstream of the overpass crest so that they can gradually reduce their speeds and avoid rear-end collisions. Therefore, accidents were selected as the primary measure of the system's effectiveness. In addition, a questionnaire was administered to obtain subjective reactions to the system.

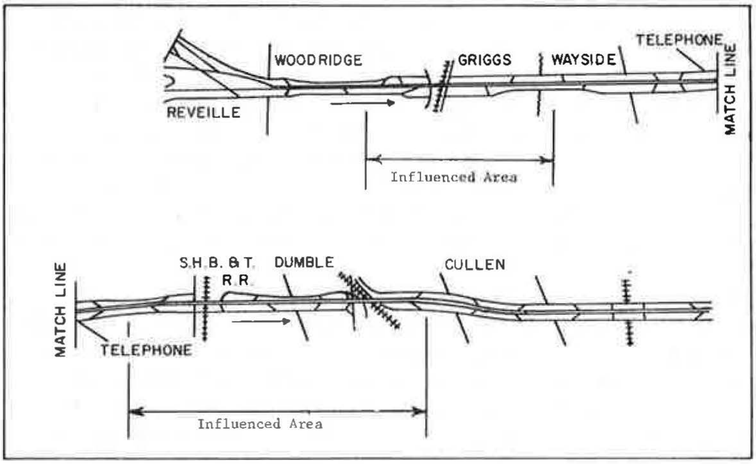
Accidents

The Houston Police Department furnished the Texas Transportation Institute daily logs of all reported accidents on the test section of the Gulf Freeway since August 12, 1971, to evaluate the use of the accident-investigation sites (3). These data provided the

Figure 1. Warning sign with flashers.

Figure 2. Flasher unit at overpass crest.

Figure 3. Freeway area influenced by warning system.



researchers with a data base to evaluate the effect of the warning system on accident experience. The warning system became operational April 3, 1972, and police accident records provided accident experience for approximately 9 months before the system was turned on. These data were compared to data from comparable dates and time periods during the first year of operation. Only accidents occurring during the operation periods (Monday through Friday from 6:30 a.m. to 6:30 p.m.) were included in the study.

Questionnaire

Studies were conducted at the 3 warning-sign sites during peak and off-peak periods. License plate numbers were recorded on hand-held tape recorders when the warning system was activated. After the tapes were transcribed, we obtained the names and addresses of the drivers within 24 hours after each field study by means of a remote terminal in the Texas Highway Department in Houston. The questionnaire sent to the drivers is shown in Figure 4.

ACCIDENTS

The number of inbound accidents before and after the warning system became operational is given in Table 1. The results show a statistically significant reduction at the 5 percent level of inbound freeway accidents. A total of 158 accidents occurred during the 9-month period before the warning system became operational, and 123 accidents occurred during the 9-month period after the system became operational. This represents a reduction of 35 accidents or 22 percent. The greatest reduction was during the morning peak period. Data for the outbound direction also are given in Table 1 and serve as a base to determine whether the changes in the inbound direction merely reflect a pattern consistent with the freeway as a whole. The results reveal that accidents in the outbound direction increased from 140 to 166 or 19 percent during the same period. The upward outbound accident trend places more significance on the inbound accident reduction.

The warning system also aimed to reduce the frequency of secondary accidents. The frequency of secondary accidents is given in Table 2. The results again reveal a statistically significant reduction in secondary collisions at the 5 percent level in the inbound direction, and, again, the secondary accidents in the outbound direction remained relatively constant. Nine secondary accidents occurred inbound before the system became operational, whereas only 1 secondary accident occurred during the same time after the warning system was operational, which is a reduction of 89 percent.

Perhaps of greater significance are the before and after comparisons to total and secondary inbound accidents within and outside the freeway sections influenced by the warning system (Fig. 3) given in Tables 3 and 4. The results show that the entire reduction in both total and secondary inbound accidents took place in the freeway section influenced by the warning system. Total accidents were reduced by 49 percent in the influenced section, whereas secondary collisions were reduced by 100 percent. There were no changes in the accident statistics in the other section of the inbound control section of the Gulf Freeway. The statistics in the outbound direction show only a slight reduction in total accidents in the same section of the freeway where warning signs influenced inbound traffic. Secondary collisions remained constant in these outbound sections.

So the warning system on the Gulf Freeway significantly reduced total and secondary accidents. That accidents in the outbound direction increased during the same time period places more significance on the utility of the warning system.

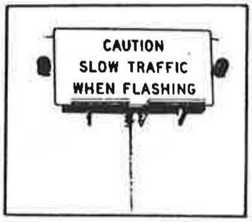
Figure 4. Questionnaire.

1. *Approximately how often do you use the inbound Gulf Freeway each week?*

   1 to 3 times per week____; 3 to 5____; 5 to 10____; over 10____

2. *Have you ever noticed the yellow sign, shown in the photograph, on the Gulf Freeway?*

   Yes____ No____



3. *Was the sign ever working when you saw it?*

   Yes____ No____

4. *About how many times have you passed it when it was working?* ____________

5. *What aspect of the sign called your attention to it?*

   ____

   ____

6. *The sign stated: "Caution Slow Traffic." How far ahead did you think it meant?*

   Over a mile____; a half-mile____; less than half mile____; less than 1 block____

7. *What speed did you think you should slow down to?*

   55____; 45____; 35____; 25____; 15____

8. *How useful was the sign to you in the actual traffic situation? (in avoiding an accident)*

   Very useful____ Moderately useful____ Limited use____ No use____

9. *Can you think of a better message that could have been on the sign?*

   Yes____ No____

   *If yes, what message?*______________________________________

10. *What did you do when you saw the sign in operation?*

    Began braking____; Slowed down gradually____; Continued at same speed, but with caution for slow traffic____; Waited until I could see the traffic ahead____

11. *To what extent was it necessary for you to slow down after you came over the overpass and saw the traffic?*

    Very little____; Moderate reduction in speed was required____; Needed to brake or change lanes____

12. *When you got over the overpass, what was the speed of the traffic ahead?*

    Same speed as before____; Moving slightly slower than before the overpass____; Moving very slowly____; Traffic was stopped in some lanes____

Table 1. Total accidents by time period.

Time Period	Before[a]	After[b]	Net Change	Percentage of Change
Inbound				
6:30 a.m. to 9:00 a.m.	60	28	-32	-53
9:00 a.m. to 4:00 p.m.	68	65	-3	-4
4:00 p.m. to 6:30 p.m.	30	30	0	0
Outbound				
6:30 a.m. to 9:00 a.m.	23	29	+6	+26
9:00 a.m. to 4:00 p.m.	68	85	+17	+25
4:00 p.m. to 6:30 p.m.	49	52	+3	+6

[a]July 12, 1971, to April 2, 1972. [b]July 12, 1972, to April 2, 1973.

Table 2. Secondary accidents by time period.

Time Period	Before[a]	After[b]	Net Change	Percentage of Change
Inbound				
6:30 a.m. to 9:00 a.m.	4	0	-4	-100
9:00 a.m. to 4:00 p.m.	5	0	-5	-100
4:00 p.m. to 6:30 p.m.	0	1	+1	
Outbound				
6:30 a.m. to 9:00 a.m.	1	1	0	0
9:00 a.m. to 4:00 p.m.	5	4	-1	-20
4:00 p.m. to 6:30 p.m.	1	3	+2	+200

[a]July 12, 1971, to April 2, 1972. [b]July 12, 1972, to April 2, 1973.

Table 3. Total accidents by freeway section.

Freeway Section	Before[a]	After[b]	Net Change	Percentage of Change
Inbound				
A[c]	72	37	-35	-49
B[d]	86	86	0	0
Outbound				
A[e]	60	55	-5	-8
B[e]	80	111	+31	+39

[a]July 12, 1971, to April 2, 1972.
[b]July 12, 1972, to April 2, 1973.
[c]Influenced by warning signs.
[d]Not influenced by warning signs.
[e]No warning signs.

Table 4. Secondary accidents by freeway section.

Freeway Section	Before[a]	After[b]	Net Change	Percentage of Change
Inbound				
A[c]	8	0	-8	-100
B[d]	1	1	0	0
Outbound				
A[e]	4	4	0	0
B[e]	3	4	+1	+33

[a]July 12, 1971, to April 2, 1972.
[b]July 12, 1972, to April 2, 1973.
[c]Influenced by warning signs.
[d]Not influenced by warning signs.
[e]No warning signs.

COST-EFFECTIVENESS ANALYSIS

Benefit Analysis

The anticipated benefits of the safety warning system were improved safety and convenience and reduction in delay time. Convenience is difficult to quantify, but it is reflected in a higher level of service resulting from fewer accidents and from the driver's confidence in conditions downstream while he or she travels at a relatively high speed.

Reduction in Accidents

The results previously discussed showed that 35 fewer accidents occurred during a 9-month period after the warning system became operational. If the rate of reduction is assumed to be consistent throughout the year, then the total would be approximately 47 fewer accidents (43 fewer peak-period accidents) during a 12-month period. Whether all 35 incidents during the 9-month period were eliminated by the warning system may be argued. But, because the accidents increased by 20 percent in the outbound section of the Gulf Freeway, it can be assumed that the warning system contributed to the bulk of the accident reduction in the inbound direction.

A convenient method that uses the chi-square test is available to determine the statistical reliability of accident reductions resulting from a safety improvement (4). Based on the chi-square test, the 22 percent reduction in total accidents, the 49 percent reduction in total accidents occurring within the influenced section, and the 100 percent reduction in secondary accidents occurring within the influenced section are all statistically significant at the 5 percent level. In other words, the accident reduction was due to the treatment rather than chance.

Burke (5) in 1970 determined costs of accidents. If we assume a 5 percent per year compounded increase, the cost per vehicle in 1972 would be \$308 for property damage accidents and \$1,857 for injury accidents.

If we assume that all the accidents analyzed involved only 2 cars and that only property damage was incurred, then the annual savings due to reduction of 47 accidents would be \$29,000.

Reduction in Delay

Goolsby (6) found that an average of 340 vehicle hours of delay results from an accident that occurs during the peak period on the Gulf Freeway; very little delay is experienced when accidents occur during the off-peak period unless the incident blocks more than 1 lane for a prolonged period of time. Pittman and Loutzenheiser (3) later estimated that use of the accident-investigation sites can reduce delay by 54 percent. Thus if the involved vehicles are removed from the freeway, estimated delay for an accident during the peak period would be 156 vehicle hours. Pittman also reported that approximately 70 percent of the accidents occurring in the study section of the Gulf Freeway are moved to accident-investigation sites or off-freeway sites for investigation and reporting. If we assume that 70 percent of the accidents that occurred during the study were removed from the freeway for investigation and reporting, then the following would reflect the estimated annual reduction in delay during the peak period due to the safety warning system:

43 accidents × 0.70 removed × 156 vehicle hours	= 4,696 vehicle hours
43 accidents × 0.30 not removed × 340 vehicle hours	= 4,386 vehicle hours
Total annual reduction in delay	= 9,082 vehicle hours

If we assume that there are 1.2 persons per vehicle and that \$4.50 per vehicle hour represents the value of time (3), then the annual monetary savings due to the reduction in delay would be \$40,850.

Cost Analysis

Gulf Freeway System

The following summarizes the initial and annual maintenance costs for the Gulf Freeway warning system:

Item	Cost
Engineering, materials, and labor costs for 3 signs	\$16,900
Estimated engineering costs for 26 detectors	2,000
Materials and labor costs for 26 detectors	21,200
Estimated annual maintenance costs	2,000

If we assume an interest rate of 7 percent for a 10-year system life expectancy, then the benefit-cost ratio (B/C) can be computed as follows:

$$B/C = \frac{AB}{(crf \times IC) + AMC}$$

where

AB = annual benefits,
crf = uniform series capital recovery factor for i = 7 percent, n = 10 years,
IC = initial capital cost, and
AMC = annual maintenance cost.

Annual benefits of the system due to reduction in delay and accidents were

$$\$29,000 + \$40,850 = \$69,850$$

Thus

$$B/C = \frac{\$69,850}{(0.1424 \times \$40,100) + \$2,000}$$

$$= \frac{\$69,850}{\$7,710}$$

$$= 9.1$$

New System

Because the warning system was added to the existing control system on the Gulf Freeway, initial cost was reduced because communications and a computer were available. Table 5 gives the cost of the same warning system if new detectors and communications

Table 5. Warning system costs for new installation.

Item	Amount (dollars)
Engineering, materials, and labor costs for 3 signs	16,900
Estimated engineering costs for 48 detectors[a]	3,000
Estimated materials and labor costs for 48 detectors[a]	39,100
Estimated costs for 1 controller (minicomputer) and associated equipment	13,400
Estimated costs for communications (telephone lines)	8,700
Estimated annual maintenance costs	3,000

[a]Estimate is based on installing 48 detectors for the 3 signs on the Gulf Freeway, including 2 sensors on each lane at each detector station. This is a requirement for using the traffic-energy control variable. It is assumed that there are 2 downstream and 1 upstream detector stations. The number of detectors would be reduced by 50 percent if the lane-occupancy-control variable is used for shock-wave detection.

had to be installed and a computer purchased. The benefit-cost ratio for the new system is computed as follows:

$$B/C = \frac{\$69,850}{(0.1424 \times \$81,100) + \$3,000}$$

$$= \frac{\$69,850}{\$14,550}$$

$$= 4.8$$

QUESTIONNAIRE

Fifteen studies were conducted at the 3 study site locations. Seven were conducted at the Griggs overpass, 3 were conducted at the South Houston, Belt and Terminal (HB&T) overpass, and 5 were conducted at the Cullen overpass. Seven were during peak periods, and 8 were during off-peak periods. All off-peak-period studies and 2 peak-period studies were conducted when an accident occurred on the freeway. Weather was clear and dry except for 3 off-peak-period studies when it was damp, drizzling, or overcast. Table 6 gives these data and indicates (a) the number of questionnaires mailed, (b) the number who returned the questionnaire forms, and (c) the number who completed the forms.

A total of 278 forms (28 percent) of the 975 mailed were returned in the 15 studies. One hundred eighteen (43 percent) of the respondents were from the 8 off-peak-period studies, and 155 (57 percent) were from the 7 peak-period studies.

Frequency of Travel on Freeway

Table 7 gives the data on detection factors and frequency of traveling on the Gulf Freeway each week.

Combined Conditions

Table 7 indicates that 106 or approximately 40 percent of all respondents drove on the freeway 5 to 10 times per week, 8 percent traveled on it more often than this, 25 percent drove on it 3 to 5 times per week, and 27 percent drove on it 1 to 3 times per week.

Peak Versus Off-Peak Conditions

During off-peak periods we expected to sample the infrequent freeway user and during peak periods we expected to sample the regular commuter. These expectations were borne out in the reports of frequency of use. Eighty-two peak-period respondents (53 percent) said that they traveled daily (5-10 times a week); only 24 (21 percent) of the off-peak-period respondents traveled daily. Fifty-two (46 percent) of the off-peak-period respondents traveled only 1 to 3 times per week or less, and only 12 percent of the peak-period respondents reported traveling this infrequently.

Detection Factors

Two hundred sixty-five (98 percent) of the 273 respondents indicated that they had noticed the sign and 236 (89 percent) said that they had seen it in operation. Ninety-

Table 6. Summary of respondent data.

Location	Period	Incident	Number of Studies	Mailed	Returned	Completed	Response Percent
Griggs	Peak	Yes	1	35	15	14	43
	Peak	No	4	356	122	118	33
	Off-peak	Yes	2	71	27	27	38
South HB&T	Peak	Yes	1	48	10	10	21
	Off-peak	Yes	2	102	28	28	27
Cullen	Peak	No	1	94	13	13	14
	Off-peak	Yes	4	269	63	63	23

Table 7. Summary of detection factors.

		Frequency of Travel on Gulf Freeway (times per week)					Number Noticing Sign			Number Noticing Sign Operating			Observed Signs Operating (number of times)				
Location (1)	Questionnaires Returned (2)	1 to 3 (3)	3 to 5 (4)	5 to 10 (5)	>10 (6)	Blank (7)	Yes (8)	No (9)	Blank (10)	Yes (11)	No (12)	Blank (13)	Always (14)	>20 (15)	<20 (16)	Never (17)	Blank (18)
Peak Period																	
Griggs	132	15	37	72	8	0	131	0	1	125	6	1	22	35	48	1	26
South HB&T	10	4	3	1	1	1	9	1	0	7	3	0	1	0	6	3	0
Cullen	13	0	4	9	0	0	12	1	0	12	0	1	2	4	5	0	2
Total	155	19	44	82	9	1	152	2	1	144	9	2	25	39	59	4	28
Percentage of total	57	12		53			99			94			20	31	47	2	—
Off-Peak Period																	
Griggs	27	10	8	5	4	0	26	1	0	22	3	2	2	3	12	1	9
South HB&T	28	13	4	6	3	2	26	1	1	25	2	1	1	4	17	1	5
Cullen	63	29	12	13	6	3	61	1	1	45	16	2	3	2	41	14	3
Total	118	52	24	24	13	5	113	3	2	92	21	5	6	9	70	16	17
Percentage of total	43	46		21			97			81			6	9	69	16	—
Peak and Off-Peak Periods Combined																	
Griggs	159	25	45	77	12	0	157	1	1	147	9	3	24	38	60	2	35
South HB&T	38	17	7	7	4	3	35	2	1	32	5	1	2	4	23	4	5
Cullen	76	29	16	22	6	3	73	2	1	57	16	3	5	6	46	14	5
Total	273	71	68	106	22	6	265	5	3	236	30	7	31	48	129	20	45
Percentage of total	—	27	25	40	8	—	98			89					57		

	Aspect Attracting Attention												
Location (1)	Flashing Light (19)	Colored Lights (20)	Lights (21)	Not Flashing (22)	Message (23)	Size (24)	Location (25)	Slow Traffic (26)	Vague (27)	Newness (28)	Colored Paint (29)	Other (30)	Blank (31)
Peak Period													
Griggs	76	15	12	4	6	7	2	2	7	3	0	1	19
South HB&T	5	0	0	0	2	0	0	0	1	0	2	0	1
Cullen	10	4	0	0	3	0	0	0	0	0	0	0	1
Total	91	19	12	4	11	7	2	2	8	3	2	1	21
Percentage of total	—[a]	—[a]	—[a]	—[a]									
Off-Peak Period													
Griggs	14	9	0	0	4	2	1	0	0	0	0	0	6
South HB&T	18	6	1	0	2	1	0	2	1	1	0	0	4
Cullen	27	14	2	1	6	7	3	2	1	0	0	0	3
Total	59	29	3	1	12	10	4	4	2	1	0	0	13
Percentage of total	—[b]	—[b]	—[b]	—[b]									
Peak and Off-Peak Periods Combined													
Griggs	90	24	12	4	10	9	3	2	7	3	0	1	25
South HB&T	23	6	1	0	4	1	0	2	2	1	2	0	5
Cullen	37	18	22	1	9	7	3	2	1	0	0	2	4
Total	150	48	45	5	23	17	6	6	10	4	2	3	34
Percentage of total	—[c]	—[c]	—[c]	—[c]	8	6							

[a]One hundred twenty-six people (81 percent) noted the sign aspects in columns 19 through 22.
[b]Ninety-two people (78 percent) noted the sign aspects in columns 19 through 22.
[c]Two hundred eighteen (80 percent) noted the sign aspects in columns 19 through 22.

four percent of the peak-period respondents had seen it in operation, whereas only 81 percent of the off-peak-period respondents had seen it in operation. Because the sign was operating at the time the drivers passed it, the negative responses could be due either to their not wishing to complete the questionnaire or to their not detecting it.

Responses to the question that asked how often the sign had been seen in operation varied greatly. Some drivers gave numerical estimates, some reported percentages of the time, and still others responded in terms of "always" or "never." Responses were classified into 5 categories as follows:

1. Always or nearly always;
2. Most of the time, 50 percent or more of the time, many times, or 20 or more instances;
3. Some of the time or less than 20 instances;
4. Never; and
5. No response.

One hundred twenty-nine (57 percent) of the respondents stated that they had seen the sign in operation some of the time. However, a peak-period versus off-peak-period comparison showed significant differences. Fifty-one percent of the peak-period and only 15 percent of the off-peak-period respondents said that it was on all or most of the time. Sixty-nine percent of the off-peak-period respondents said that it was on some of the time. Only 47 percent of the peak-period respondents reported seeing it working only occasionally.

Twenty-eight of the peak-period respondents and 17 of the off-peak-period respondents left the question blank. This question was the one most frequently not answered.

Two hundred eighteen respondents (80 percent) said that flashing lights, yellow lights, lights only, or lights not flashing attracted their attention. All of these write-in responses were judged to be indications that the flashing beacons had attracted drivers' attention to the signs. Of the peak-period respondents, 81 percent noted the lights; of the off-peak-period respondents, 78 percent noted them. The next most frequently mentioned aspects were the message, CAUTION (8 percent), and the size of the sign (6 percent). Other comments included references to the visibility or appearance of the sign, its location, slow traffic in the area, newness of the sign, and color of the sign. Twelve percent left the question blank. No appreciable differences were found between peak-period and off-peak-period respondents or among various sign locations.

Interpreting the Message

Table 8 gives the respondents' answers on interpreting the meaning of the displayed message, overall evaluation of the sign's usefulness, actions taken in response to the sign, and relevance of the message to what was later observed about traffic conditions downstream of the overpass.

Distance Meaning

Eighty-nine percent of the respondents (91 percent of the peak-period and 87 percent of the off-peak-period respondents) expected the slowdown to occur from 1 block to 0.5 mile (0.805 km) away, whereas 11 percent expected the congestion to occur 1 mile (1.609 km) or more ahead. Almost half the respondents felt that the message, SLOW TRAFFIC AHEAD, referred to a distance of less than 0.5 mile (0.805 km) but more than 1 block away. Very little difference between the peak-period and non-peak-period respondents was reported.

Table 8. Summary of interpretation, evaluation, and response factors.

Location	Questionnaires Returned	Distance Meaning: 1 Mile	0.5 Mile	<0.5 Mile	<Block	Blank	Speed Meaning: 55 mph	45 mph	35 mph	25 mph	15 mph	Blank	Amount of Usefulness: Great	Fair	Limited	None	Blank
Peak Period																	
Griggs	132	11	24	58	32	7	0	14	40	47	22	13	41	32	31	25	3
South HB&T	10	1	1	5	1	2	0	2	3	0	0	5	3	5	0	2	0
Cullen	13	1	3	7	1	1	0	0	3	4	4	2	5	4	2	1	1
Total	155	13	28	70	34	10	0	16	46	51	26	20	49	41	33	28	4
Percentage of total	57	9	19	48	24	—	0	11	33	37	19	—	32	27	22	19	—
Off-Peak Period																	
Griggs	27	3	3	8	7	6	0	3	6	7	5	6	13	6	1	2	5
South HB&T	28	3	2	19	1	3	0	3	11	10	2	3	9	14	1	1	3
Cullen	63	8	14	23	14	4	0	12	26	16	4	5	24	19	8	7	5
Total	118	14	19	50	22	13	0	18	43	33	11	14	46	39	10	10	13
Percentage of total	43	13	18	48	21	—	0	17	41	32	10	—	44	38	9	9	—
Peak and Off-Peak Periods Combined																	
Griggs	159	14	27	66	39	13	0	17	46	54	27	19	54	38	32	27	8
South HB&T	38	4	3	24	2	5	0	5	14	10	2	8	12	19	1	3	3
Cullen	76	9	17	30	15	5	0	12	29	20	8	7	29	23	10	8	6
Total	273	27	47	120	56	23	0	34	89	84	37	34	95	80	43	38	17
Percentage of total	—	11	19	48	22	—	0	14	36	35	15	—	37	31	17	15	—

Location	Other Message Ideas	First Action: Braked	Slowed	Continued With Caution	Waited	Blank	Secondary Action: Little	Reduced Speed	Braked	Blank	Credibility: Same Speed	Slightly Slower	Very Slow	Stopped	Blank
Peak Period															
Griggs	18	6	66	48	5	7	35	70	19	8	19	54	38	27	5
South HB&T	3	0	4	3	0	3	1	7	0	2	1	5	2	0	2
Cullen	2	0	7	3	2	1	3	8	2	0	3	5	6	6	1
Total	23	6	77	54	7	11	39	85	21	10	23	64	46	33	8
Percentage of total	—	4	53	38	5	—	27	59	14	—	14	39	27	20	—
Off-Peak Period															
Griggs	4	3	16	4	1	7	0	16	5	7	1	12	3	5	7
South HB&T	4	4	18	2	1	3	4	16	3	4	1	10	7	7	4
Cullen	13	7	29	14	2	11	10	30	14	10	5	24	15	12	8
Total	21	14	63	20	4	21	14	62	22	21	7	46	25	24	19
Percentage of total	18	14	62	20	4	—	14	63	23	—	6	45	25	24	—
Peak and Off-Peak Periods Combined															
Griggs	22	9	82	52	6	14	35	86	24	15	20	66	41	32	12
South HB&T	7	4	22	5	1	6	5	23	3	6	2	15	9	7	6
Cullen	15	7	36	17	4	12	13	38	16	10	8	29	21	18	9
Total	44	20	140	74	11	32	53	147	43	31	30	110	71	57	27
Percentage of total	16	8	57	30	5	—	22	61	18	—	11	41	27	21	—

Note: 1 mile = 1.609 km.

Speed Meaning

The message on the sign implied that traffic should slow down to some safe speed. Slightly over a third of the respondents felt that this speed was 35 mph (56.3 km/h), and another third felt that it was 25 mph (40.2 km/h). Those driving during the peak period felt that the sign meant a lower speed than did those driving during the off-peak period. Fifty-six percent of the peak-period group selected either 25 or 15 mph (40.2 or 24.1 km/h) compared to 42 percent of the off-peak-period group. Also about 6 percent more of the off-peak-period respondents felt that the message implied that 45 mph (72.4 km/h) was the safe speed. We anticipated this finding because of the higher traveling speeds during off-peak conditions.

No one selected 55 mph (88.5 km/h), which was the speed limit itself; selecting it would have implied that the driver was traveling faster than the legal limit.

Usefulness

Sixty-eight percent of the respondents stated that the sign was either very or moderately useful to them. However, there were significant differences of opinion between peak-period and off-peak-period drivers on its usefulness. Eighty-two percent of the off-peak-period drivers who responded stated it was useful, and only 59 percent of the peak-period drivers felt that it was useful. The higher percentage of negative responses in the peak-period group was borne out by write-in comments on the forms that the message was not informative when prevailing traffic conditions were already stop-and-go.

Responses to the Message

Respondents were asked 2 questions. The first related to their immediate reaction on seeing the sign, and the second related to their need for additional reduction in speed after they passed the crest and could see the actual state of traffic.

Immediate Reaction

Fifty-seven percent of the respondents reported that they slowed down gradually when they saw the sign; 30 percent stated that they would continue at the same speed with caution. Only 8 percent said that they would brake, and 5 percent said that they would brake, and 5 percent said that they would wait to see the traffic before doing anything.

A comparison between peak-period and off-peak-period respondents revealed that 62 percent of the off-peak-period respondents said that they slowed down gradually; only 53 percent of the peak-period respondents selected this response. This difference might be interpreted in terms of vehicle speeds and the opportunity to slow down further.

Thirty-eight percent of peak-period respondents said that they would continue with caution; only 20 percent of the off-peak-period respondents selected this response. Again, the off-peak-period drivers had greater opportunity to slow down so fewer drivers selected this response. Peak-period drivers were somewhat more compelled to drive at the prevailing traffic speed; hence more drivers continued cautiously at the same speed.

Secondary Action

Sixty-one percent of the respondents indicated that they needed to reduce their speed moderately after they came over the overpass and saw the traffic. Peak-period and off-peak-period drivers responded to the same degree. Ideally this would not have been necessary. The typical reaction was not only to slow down at the sign but also to wait for some visual feedback from the traffic ahead before adjusting the speed to the prevailing traffic flow. This response would be satisfactory except when the stoppage wave was immediately downstream of the crest—a possibility that only 22 percent of the drivers anticipated. Twenty-two percent of the respondents indicated that they had to do very little in adjusting their speed after passing the crest. Twice as many peak-period as off-peak-period respondents indicated that they did little adjusting. Again this may be due to the comparative lack of opportunity to reduce speed.

Eighteen percent of the respondents admitted that they had needed to brake or change lanes (an admission that the sign was truthful) but that they had not responded appropriately to the message. However, this does not mean that they would respond inappropriately in future encounters. As we expected, more off-peak-period drivers needed to brake than did peak-period drivers.

Message Credibility

The last question measured the respondents' interpretation of the validity of the system and the credibility of the message, SLOW TRAFFIC. Respondents were asked to select the actual state of traffic that they had encountered. A statement that traffic downstream was traveling at the same speed as it was upstream would be tantamount to stating that the system was not working. Only 11 percent of all respondents selected this response (14 percent of the peak-period and 6 percent of the off-peak-period respondents). This suggests that off-peak-period drivers, who generally were not exposed to the sign under stop-and-go conditions, found the message more credible.

Forty-one percent of all respondents said that the traffic was slightly slower; 27 percent reported that it was very slow; and 21 percent reported stoppages. Peak-period and off-peak-period percentages were similar.

FINDINGS AND RECOMMENDATIONS

The results of the study suggest that the warning system on the inbound Gulf Freeway is a cost-effective system for alerting approaching motorists to stoppages on the freeway. The warning system significantly reduced total and secondary accidents on the freeway. The following specific findings can be drawn from the results of this research:

1. The warning system on the Gulf Freeway resulted in an estimated annual reduction of approximately 47 accidents and 9,082 vehicle hours of delay. The benefit-cost ratio was estimated to be 9.1.
2. Because the warning system was integrated with the existing control system on the Gulf Freeway, considerable initial cost was avoided. (An analysis of a new system that assumed that there was no available hardware resulted in a benefit-cost ratio of 4.8.)
3. Studies of accidents for 9-month periods before and after the warning system began operation revealed that accidents were reduced from 72 to 37 or 49 percent in the sections of the inbound Gulf Freeway influenced by the warning system, and accidents in comparable outbound sections were reduced from 60 to only 55 or 5 percent. The greatest inbound accident reduction occurred during the morning peak period. There was a 100 percent reduction in secondary accidents (8 before, 0 after) in the inbound freeway section influenced by the warning system. Essentially no change in secondary accidents occurred in the other inbound or outbound freeway sections.
4. The results of the questionnaire study indicated that the motorists observing the sign in operation believed that the sign was useful and readily noticed and that the message was generally understandable. The respondents reacted to it appropriately and confirmed that the message displayed was verified later by traffic conditions. The sign was especially effective and accepted during the off-peak period when motorists were traveling at higher speeds and approached an incident not visible to them.
5. The greatest skepticism regarding the usefulness of the sign came from the

peak-period respondents. Fifty-one percent reported seeing the sign in operation all or most of the time compared to 15 percent of the off-peak-period respondents. Although both groups reported that the sign was useful, 9 percent more off-peak-period than peak-period drivers said that they would slow down gradually; 18 percent more peak-period than off-peak-period drivers said that they would continue with caution when they saw the sign, presumably because they were not able to slow down very much. Twice as many peak-period as off-peak-period drivers said that they needed to do very little when they saw the traffic; again this suggests that there was not need for action because of the prevailing traffic speed. Peak-period drivers also criticized the fact that the sign was on most of the time and presented information that was obvious to stop-and-go drivers.

There was a contradiction between the accident study results and questionnaire responses. Drivers, particularly those at the Griggs location, complained that the sign was activated most of the time during the peak periods. The statistics, however, showed a large reduction in total and secondary accidents during the peak periods. These results suggest that the warning system should be operated during the peak period but that the sign should be turned off as quickly as possible when the shock wave passes over the crest. This can be accomplished by placing the upstream sensors as close as possible to the structure.

The results verify that the flashing beacons were effective and provided excellent target value. And, although it may be desirable to state an indicated safe speed and the distance ahead to which the sign applies, a sufficiently large percentage of drivers interpreted the distance to be 0.5 mile (0.805 km) or less. They also felt that the sign implied a safe speed of 15 to 35 mph (24.1 to 56.3 km/h) except when the traffic was actually stopped immediately over the crest. The sign would be useful within the constraints of a fixed message.

ACKNOWLEDGMENTS

We wish to acknowledge the assistance of Richard Fullerton of the Texas Transportation Institute who coordinated the questionnaire studies.

This paper discusses 1 phase of a research project, Development of Urban Traffic Management and Control Systems, that is being conducted by the Texas Transportation Institute and the Texas Highway Department in cooperation with the Federal Highway Administration. The contents of this paper reflect the views of the authors who are responsible for the facts and the accuracy of the data presented. The contents do not reflect the official views or policies of the Federal Highway Administration.

REFERENCES

1. C. L. Dudek and R. G. Biggs. Design of a Safety Warning System Prototype for the Gulf Freeway. Texas Transportation Institute, Texas A&M Univ., Research Rept. 165-4, May 1972.
2. C. L. Dudek and C. J. Messer. Detecting Stoppage Waves for Freeway Control. Highway Research Record 469, 1973, pp. 1-15.
3. M. A. Pittman and R. C. Loutzenheiser. A Study of Accident Investigation Sites on the Gulf Freeway. Texas Transportation Institute, Texas A&M Univ., Research Rept. 165-1, Aug. 1972.
4. R. M. Michaels. Two Simple Techniques for Determining the Significance of Accident-Reducing Measures. Traffic Engineering, Vol. 36, Sept. 1966.
5. D. Burke. Highway Accident Costs and Rates in Texas. Texas Transportation Institute, Texas A&M Univ., Research Rept. 144-1F, 1970.
6. M. E. Goolsby. Accident Reporting and Clearance Procedures on the Gulf Freeway. Texas Transportation Institute, Texas A&M Univ., Research Rept. 139-1, Sept. 1969.

CORQ—A MODEL FOR PREDICTING FLOWS AND QUEUES IN A ROAD CORRIDOR

Sam Yagar, University of Waterloo, Ontario

A procedure has been developed for predicting the self-assignment of time-varying traffic demands in a network. The procedure's computer program, CORQ, has been used to validate and apply the model in a real corridor. It is intended as a tool to enable the traffic analyst to assess the systemwide effects of any traffic-control strategies proposed for a network as long as the total system's demands remain invariant or at least have a predictable response to the controls. The model has been specialized to give detailed treatment to the critical elements of a corridor that affect traffic flow, capacity, queuing, and delays. It can be used for a form of microanalysis of areas that are about 500 blocks large. For these cases it considers only the major intersections, freeway interchanges, and their surface-street links but gives them a detailed treatment. It can be used for much larger areas if only the freeway network needs to be modeled. Time-varying traffic controls can be simulated. CORQ also can serve as a partial optimization technique by selecting metering rates that fully use the capacity of a merge without queuing on the freeway. CORQ is intended for use in estimating quantitatively the effects of various types of traffic-control strategies before a commitment to any specific control schemes and installation of specialized hardware. It can serve as a traffic-management game, and it has been used in training students in the design of traffic-engineering and traffic-control schemes including ramp closure, ramp metering, restriping, and altering traffic-signal splits.

•THE DAYS of eminent domain and resultant easy financing and justification of roads are ending. The capacities of our urban roadway corridors are leveling off while demands continue to grow. Traffic engineers, often prodded by citizens and politicians, are looking increasingly into more efficient use of existing roadways.

May (1) and others recognized the potential benefits and reduction in overall delay that might be achieved in corridor use through closure or metering of freeway ramps. This concept has been applied with varying degrees of success. The initial Chicago application resulted in improved freeway operation at the expense of surface-street operation. There was little or no net improvement (1). Since then, freeway-control experts have developed certain subjective rules based on both theory and experience that have improved the probability and average level of success.

Nevertheless there remained a residual need for an evaluative tool that could be used to weigh various types of freeway-control strategies or exact control schemes or both before a commitment to their application in the field. Such a model is described in this paper.

REQUIREMENTS OF AN EVALUATIVE MODEL

Data

To minimize any added requirements for data, the model should attempt to use the type of data that are being collected in freeway-control studies. These generally 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. In addition, much data collection is repeated in afterstudies to determine operating conditions with the controls implemented so that the controls may be evaluated. These types of information are generally used to form control strategies.

Simplicity

The model should be reasonably easy to understand and apply.

Precision

Accurate representation of the time variation in traffic demands is especially critical for peak periods during which temporary high-demand levels can lead to oversaturations and cause queuing and delays. Small oversaturations can produce queues that often persist for the entire peak period until they are relieved generally by a return to levels of demand that are below capacity. A microanalysis is needed to emulate the peak-period traffic operation of critical sections in sufficient detail so that even small oversaturations can be detected and the effects of resultant bottlenecks can be quantified accurately.

Sensitivity

The model should be able to predict driver response to each type of traffic control. It should be able to predict not only how the directly controlled traffic will respond but also how the second-order effects are on the paths of indirectly affected users who respond to the actions of those directly controlled. The traffic engineer is more likely to be able to predict the former. The interactions in the latter can become too complicated or at least too cumbersome for repeated application with each of the various control schemes that he or she may wish to evaluate. High-speed computer simulation is ideally suited to simulating repetitive cumbersome calculations if the operation can be modeled.

DESCRIPTION OF THE CORQ MODEL

A methodology called CORQ has been developed for modeling the operation of a corridor (a network with a dominant direction whose flows are of interest). It is felt that the method satisfies all of the previously mentioned requirements. Its sensitivity in modeling the effects of traffic controls is illustrated elsewhere (2).

CORQ gives detailed treatment of the critical elements of a corridor in terms of traffic flow, capacity, queuing, and delays. It is related to another specialized technique called FREQ, which emphasizes the modeling of freeway queues (3).

CORQ is a form of microassignment technique, but it is different from most of the existing techniques. For instance, it is completely different from the Brown and Scott technique (4) although both can be used for microanalysis 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. Most others do not allow for time-varying demand at all.

For accuracy, modeling the assignment of peak-period traffic to corridors rather than to general networks has been emphasized. Although the method could be applied to general networks, it was felt that there was more of a need for a predictive tool for microanalysis of corridors to deal with peak-period problems. Detailed discussion of ways of transforming the modeling procedure and methodology for more general application might tend to obscure the description of the main thrust of the work described herein.

The modeling procedure used by the methodology follows Yagar's basic outline (5) except that certain assumptions have been relaxed and additional capabilities added (6). The basic method still divides 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 min. 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.

Queues of vehicles were treated initially as if they were stored at the upstream node of the link for which they were queued (5). The queuing cost was added to the travel cost to obtain the total cost of using that link. Yagar has added provision for more accurately modeling the effects of queue spillbacks (6). In this way the effects on other vehicles and upstream capacities are better represented. The cost of queuing is dynamically approximated as directly proportional to the size of the queue and inversely proportional to the rate at which its contents are served. The model now associates queue cost with the link on which the queue occurs rather than with the bottleneck link that causes the queuing.

Provision has been made for exogenously 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).

Yagar's basic model (5) is based on an incremental assignment procedure. The main disadvantage of incremental techniques is that they can prematurely assign demands to ultimately incorrect links. A later Yagar technique (7) is used that reduces the amount of premature assignment by iterating on successive incremental solutions. Each iteration weighs in estimates of the equilibrium link-travel costs on the basis of the results of the previous iteration.

Another major problem addressed by the CORQ model relates to preestimating equilibrium capacities that depend in turn on equilibrium flows. This problem has received little attention in the literature, but it is important to traffic assignment, especially to dynamic assignment. Because delay is very sensitive to the difference between demand and capacity, both demand and capacity must be known accurately for one to reasonably estimate delay. That time produces great variations in demands is accepted. Less attention has been given to the fact that capacity also can vary as flows vary. Capacity variations occur mainly in weave sections and at merges. Although an appropriate method for estimating weave capacities for our purposes does not yet exist, the problem does, and it is discussed in another report by Yagar (8). Merging phenomena also are discussed at length in this report in which a method is described for dynamically estimating the merge capacities that will prevail when demand has been assigned to the network. For the purposes of this paper it is sufficient to note that the capacities of the approaches to a merge depend on each other's flows. The model attempts to determine these capacities along with the flows. This is especially important at freeway merges, where the capacity is shared by the main-line and on-ramp vehicles. At a given merge each approach will be able to discharge a certain minimum number of vehicles, called its flow entitlement, regardless of the demand at the competing approach. If one of the approaches does not need its full entitlement, the excess reverts to the competing approach. CORQ attempts to model this phenomenon of mutually dependent merge capacities with a capacity-sharing routine described by Yagar (8).

With its capacity-sharing routine, CORQ can serve as a partial optimization technique. The merge-sharing routine can be set to allow all main-line traffic into the merge and dynamically adjust the ramp-metering rate so that the ramp flow equals the merge capacity minus the main-line flow. The simulation results would show a metering rate that fully used the merge and no queuing on the freeway. This corresponds to traffic-responsive metering with no minimum metering rate.

The evolution of the methodology from Yagar's skeleton model (5) to the present CORQ model is given in Table 1. The basic model is characterized by:

1. Some double accounting in estimated cost of travel within a queue,
2. Use of only preestimated capacities for approaches to merges (no dynamic estimating),
3. Use of straight incremental assignment with no preestimate of equilibrium costs to find shortest paths, and
4. No consideration of upstream effects of physical backup of queues.

The sequence of steps in Table 1 indicates the additions made that hopefully will aid the reader in understanding the properties of the model. 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 (not yet in CORQ).
   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.

If varying the entitlements from iteration to iteration is desired, estimate merge capacity entitlements for the next iteration on the basis of demands and ultimate entitlements. A more detailed description of the logic and a listing of the computer program and instructions for its use are given elsewhere (8).

The CORQ model was tested on the Ottawa Queensway corridor (9). 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 (2). 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. These are discussed further elsewhere (2).

The CORQ model resembles a traffic management game as well as a simulation because it assigns users to shortest-time paths. It has been used in training students in

Table 1. Major evolutionary changes in the development of CORQ.

Step	Name	Changes
0	PROG 0	Some provision for estimating the effects of sharing capacity at merges
		Emulation of the effects of queue spillbacks
2	PROG 2	2 iterations in each time slice (First iteration is used to estimate the equilibrium unit costs on the links, which are then weighed into the costs used in the second iteration. The first iteration provides for weighing in the final costs of the previous time slice.)
8	PROG 8	Can use any even number of iterations and specify upper and lower bounds on this number for consistency in the face of oscillations
		Can specify how many iterations allow sharing of merge capacity Before sharing of merge capacity, the approach that lends capacity reserves an amount to reflect its queue at the end of the previous iteration Can update merge capacity entitlements for each iteration in line and provide more recent estimates of demands on the approaches
9	PROG 9	Double accounting eliminated
		Improved estimation of the composite cost of travel in a queue

the design of traffic-engineering and traffic-control schemes including combinations of ramp closure, ramp metering, restriping, and changes in traffic-signal splits.

ASSUMPTIONS AND APPROXIMATIONS INHERENT IN THE METHODOLOGY

Assumptions and Implications

Queue Dissipation

A queue that dissipates in a certain time slice is assumed to decrease at a constant rate over the entire length of that time slice and thus disappear at the end of the slice. This is illustrated by the dotted line in time slice n+j of Figure 1. This assumption is really effectively an approximation to the total queue time on a link. The error of approximation is the area between the full line farthest to the right and the dotted line. An upper bound for this error is $S/2\ (q_{n+j-1})$. Unless there is a drastic drop in demand in time slice n+j to dissipate a large queue, q_{n+j-1}, combined with a long time-slice length, S, this error will not be large.

Queue Evolution

A queue that exists on a certain link at the end of a time slice is assumed to have been taken out of the network and fed back in as new demand originating at the downstream end of that link. This new demand is fed in at a constant rate over the duration of the following time slice. This causes the queue evolution of Figure 1 to be approximated by the dashed trajectory. Using the assumption that the queue dissipated at the end of time slice n+j, one finds that the total queue time as approximated by the dashed curve is half the actual queue time. This can be proved by using pairs of triangles with equal area in Figure 1. Total queue time obtained by the outlined procedure has been doubled in the CORQ program to correct for this.

Driver's Knowledge of Travel Times

The model assumes that the driver knows the unit travel times of all the links for the present time slice but not for the next time slice. This means that the present best path can be chosen for the driver, but, if that path leads to a queue, he or she will select the remainder of the path based on new information when he or she is ready to leave the queue. Because relative conditions on the competing paths do not change drastically from one time slice to the next, this assumption is generally harmless.

Unlimited Queue Storage Capacity on Surface Streets

It is assumed that queues will not spill back through major intersections on surface streets. This is reasonable because the spacing of major intersections is generally quite large; it approximates the spacing of urban freeway interchanges. However, there is provision in the model for queues on freeways and ramps to extend back onto freeway, ramp, and surface-street links.

Approximations and Effects

Constant Turning Equivalents

A given type of turning movement at a given intersection is assumed to have a constant through-flow equivalent in terms of its effect on the intensity of flow at the intersection; that is, the intensity of flow at the intersection is independent of the number of such movements. This is approximately true for the small ranges of flows that one might expect to encounter at intersections in peak periods. Flow equivalents can be estimated in these small flow ranges. For example the through-flow equivalent for a left turn on a given link might be about 1.3 in an off-peak period and about 2.5 in the peak period.

Flow-Cost Relationship

The relation between unit travel time and flow for each of the links is approximated by pieced constant components. This technique replaces a link by a number of sublinks in parallel, each of which has a constant unit cost as shown in Figure 2. Yagar (10) tested this type of approximation and found the error to be small.

Unit Queue Cost

The unit cost that a user pays in waiting for a queue of vehicles, q, to be served is proportional to q as represented by the straight line shown in Figure 3. If the queue has a size, CSQ, that takes a time slice, S, to serve, the unit queue cost is S. This straight line is approximated by constant components that have capacity limits equal to 2 percent of CSQ and cost increments equal to $0.02 \times S$ as shown. For example, if a time slice is 15 min, and 1,000 vehicles can be served in a time slice, then pieced constant components would have capacities of 20 vehicles and unit time increments of 18 sec.

This level of approximation has been chosen as a compromise between accuracy and computer time. The unit queue cost is updated after each increment in the assignment even if the capacity at that cost has not been exhausted. That is, the cost is increased to a level that will allow an additional $0.2 \times$ CSQ units of queued vehicles. This is equivalent to sliding the pieced constant curve in Figure 3 along the straight line. It is done to avoid excessive and unnecessary iterations. If the previous increment added 5 vehicles to the queue, unit cost would be increased by $5/20 \times 18 = 4.5$ sec. The capacity at the new cost is 20, not 15.

Figure 1. Evolution of a queue over time.

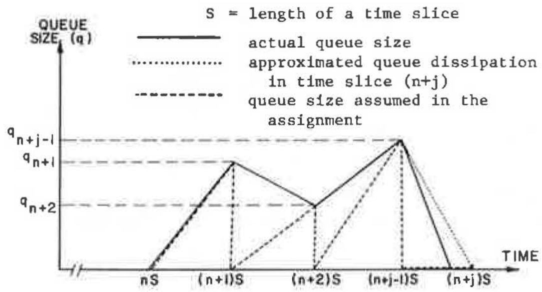


Figure 2. Unit cost versus flow.

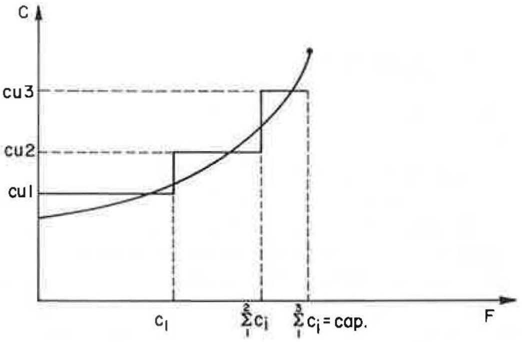


Figure 3. Linear unit cost of queuing.

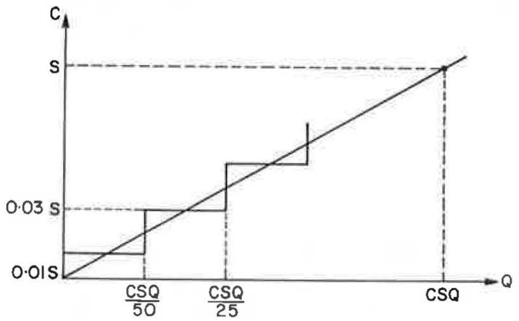


Figure 4. Intersection approach and its movements.

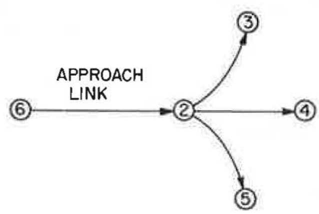


Figure 5. Concentration point that avoids illogical paths for exogenous flows by 2 nodes.

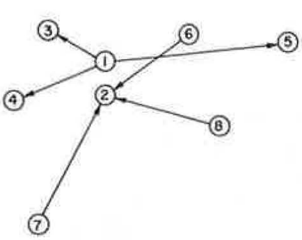


Figure 6. Dummy links for modeling capacity for sharing a merge.

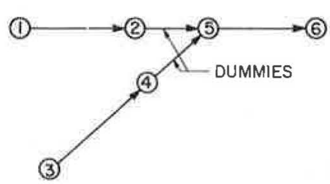
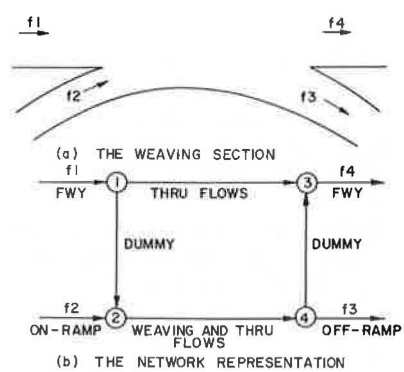


Figure 7. Simple weaving section and its network representation.



MODELING A CORRIDOR

Basic Framework

The framework for modeling a roadway corridor is based on that used by Yagar (5) together with some extensions. The roadways are represented by links that begin and terminate at nodes. The latter should occur at points where demand or flow characteristics such as capacity or both can change. This would be consistent with the procedure recommended in the Highway Capacity Manual (11) for dividing a roadway into links that are homogeneous sections. However, to minimize the size of the network used to represent the corridor, a portion of roadway is generally taken as a single link if all of that portion has the same flow. Its capacity is estimated at its point of minimum capacity. If the section's flow can change where it meets another link or has exogenous demands, it should be divided into at least 2 links by nodes at these points. If its flow can change significantly in spite of homogeneity of demand along its length, it is not a physically homogeneous section and may have to be approximated by more than 1 link with differing characteristics. The capacity of a link is treated by the model as that link's ability to absorb vehicles. A link can discharge all of the vehicles that it absorbs provided that these can be absorbed by the downstream link.

Specialized Modeling for Specific Sections

At-Grade Intersections

An approach to an at-grade intersection is modeled as shown in Figure 4. In this example movements are represented by the links 2, 3; 2, 4; and 2, 5; the approach to the intersection is represented by link 6, 2, which is a dummy link representing a section of zero length. It is used as a means of combining the magnitudes of the individual movements into a weighted total that represents an equivalent total through flow. In this way one can represent the mutual effects of the 3 types of movements on one another in spite of the fact that they generally have different unit effects on the level of traffic intensity at the intersection. Through-flow equivalents have been used extensively by Miller (12).

Avoidance of Illogical Paths

Avoiding illogical paths can be accomplished in various ways, depending on the situation, and is a matter of individual choice. One method is shown elsewhere (5, Fig. 6). Illogical paths also can be created when a single aggregation point node is used for exogenous flows both into and out of the network. In Figure 5, if nodes 3 through 8 were all joined to a single aggregation point the routine for building shortest-path trees would be able to use these exogenous links for through flows. This can be overcome by representing the concentration point by 2 nodes such as nodes 1 and 2 in Figure 5. There is no illegal shortcut through either node 1 or 2 in Figure 5 because links feed only into node 2 and out of node 1.

Merge Sections

Representation of the merging into a single roadway of 2 upstream roadways that share a total downstream capacity is shown in Figure 6. The 2 merging roadway sections are represented by links 1, 2 and 3, 4 respectively, and the downstream section is represented by link 5, 6. In addition to these, dummy links 2, 5 and 4, 5 have been inserted as shown. The dummies hold the key to modeling the sharing of total merge capacity. They represent arbitrarily short sections at the downstream ends of their respective merge links. They are given certain capacities to accept vehicles; these capacities then are used to regulate the capacities of the merging roadways to discharge vehicles onto the link downstream of the merge point. By manipulating the capacities of these dummy links, the analyst can control the model's sharing of the merge capacity to some extent. Some methods for controlling merge sharing and their implications are discussed elsewhere (8).

Weave Sections

A simple weave section and a suggested form of network representation in terms of nodes and links are shown in Figure 7a and 7b respectively. The representation in Figure 7b allows one to treat the weave and nonweave sections separately on links 2, 4 and 1, 3 respectively. Links 1, 2 and 4, 3 are dummy links. The on-ramp flow, f_2, must use links 2, 4 and 4, 3 and the off-ramp flow, f_3, must use links 1, 2 and 2, 4. Any on-off flows included in f_2 and f_3 must use link 2, 4. All of these must use the weaving section 2, 4. The through flows have the choice of using the weave section via links 1, 2; 2, 4; and 4, 3 or the nonweave section 1, 3. Their individual choices would depend on the relative conditions of the paths. This is consistent with actual operation in which the right lane is used by through vehicles when it operates as well as the through section does but is avoided by them when it is more congested.

The capacity of the weave section and the effective number of lanes used by weaving vehicles can be estimated for a given set of weave flows. The capacity of the nonweave section can be estimated from the effective number of lanes not used by weaving vehicles. The Highway Capacity Manual (11) deals with capacity of weave sections and equivalent land use of weave flows. However, the more recent work of Pignataro (13) gives weaving a more complete treatment.

SUMMARY

A model for predicting the flows and queues in a road corridor has been developed. Its computer program, CORQ, has been programmed in FORTRAN IV. It combines the following techniques:

1. Dynamic traffic assignment of time-varying demands employing queuing when the best path has a queue on it;
2. Emulation of queue spillback and its upstream effects;
3. Provision for altering network characteristics during the simulation period to allow for control strategies such as time-varying metering rates for on-ramps;
4. A traffic assignment technique that combines iterative and incremental techniques; and
5. Routines for determining the mutually dependent capacities on the approaches to a merge, for any of the following: (a) uncontrolled merge, (b) fixed metering rate for 1 approach, and (c) traffic-responsive metering.

CORQ is intended as a tool to enable the traffic analyst to assess the systemwide effects of applying traffic controls in a network as long as the total system's demands remain invariant or at least have a predictable response. It has been specialized to give detailed treatment to the critical elements of a corridor in terms of traffic flow, capacity, queuing, and delays. It can be used as a form of microanalysis of areas about 500 blocks large. For these cases it considers only the major intersections, freeway interchanges, and their surface-street links, but it gives them a detailed treatment. It also can be used for much larger areas if only the freeway network needs to be modeled. Time-varying traffic controls can be simulated. CORQ also can serve as a partial optimization technique because it can estimate main-line and on-ramp flows for any given type of control strategy by which the merge is fully used, and it will not create queuing on the freeway.

Although CORQ cannot determine exact optimal metering rates, it can determine the best possible types of control schemes. Determining exact optimal metering rates is difficult because all tests have to be based on collected data, which are only estimates of demands. The value of the CORQ model is in its estimating the effects of various types of proposed schemes on total travel time before a commitment is made to a general control scheme and finances are committed to the installation of hardware. The control hardware can be fine-tuned to optimal rates corresponding to the conditions that exist when it is in use. Determination of an appropriate type of control scheme is not sensitive to reasonable approximations in the data. CORQ also can serve as a traffic management game and has been used in training students in the design of traffic-engineering and traffic-control schemes including ramp closure, ramp metering, restriping, and altering traffic-signal splits.

AREAS FOR FURTHER RESEARCH AND DEVELOPMENT

It is felt that control strategies are not overly sensitive to exact O-D patterns except for the O-D patterns of users that might be significantly affected by control measures. It might therefore be worthwhile to find a method for manufacturing a simple set of O-D demands that would serve for testing traffic-control strategies. This might involve representing control-sensitive users by actual O-D patterns and filling in other O-D patterns so that CORQ can reproduce counted flows. In this way one could simultaneously develop the O-D matrices and calibrate the model to a given network.

CORQ also could be used to test the effects of temporal changes in demands by schemes such as staggered work hours.

ACKNOWLEDGMENTS

The development of the methodology was financed by a contract with the Transportation Development Agency of the Canada Ministry of Transport, a grant from the National Research Council of Canada, and a Canada Department of National Defence Fellowship. The testing and application of the CORQ model were sponsored by a contract with the Ontario Ministry of Transportation and Communications. I am grateful to the following individuals for some stimulating and useful discussions related to portions of the methodology and potential application of the model: W. Ross Blunden, Ken Crowley, Michael Dunne, Wolf Homburger, Roy Loutzenheiser, Adolf D. May, W. Richard McCasland, Joseph McDermott, Bill McShane, Karl Moskowitz, Louis Pignataro, and Roger Roess.

REFERENCES

1. A. D. May. Experimentation With Manual and Automatic Ramp Control. Highway Research Record 59, 1964, pp. 9-38.
2. S. Yagar. Measures of the Sensitivity and Effectiveness of CORQ. To be published in a future Transportation Research Record.
3. Y. Makigami and W. L. Woodie. Freeway Travel Time Evaluation Technique. Highway Research Record 321, 1970, pp. 33-45.
4. G. H. J. Brown and R. S. Scott. Micro-Assignment: A New Tool for Small-Area Planning. Highway Research Record 322, 1970, pp. 149-161.
5. S. Yagar. Dynamic Traffic Assignment by Individual Path Minimization and Queueing. Transp. Res., Vol. 5, 1971, pp. 179-196.
6. S. Yagar. Applications of Traffic Flow Theory in Modeling Networks. To be published in a future Transportation Research Record.

7. S. Yagar. Emulation of Dynamic Equilibrium in Traffic Networks. International Symposium on Traffic Equilibrium Methods, Montreal, Nov. 1974.
8. S. Yagar. Dynamic Assignment of Time-Varying Demands to Time-Dependent Networks. Transportation Development Agency, Canada Ministry of Transport, 1974.
9. S. Yagar. An Analysis of the Morning Peak Traffic Problems in the Eastbound Corridor Serving Ottawa. Ontario Ministry of Transportation and Communications, 1973.
10. S. Yagar. Analysis of the Peak Period Travel in a Freeway-Arterial Corridor. University of California, Berkeley, PhD dissertation, March 1970.
11. Highway Capacity Manual. HRB Spec. Rept. 87, 1965.
12. A. J. Miller. Australian Road Capacity Guide. Australia Road Research Board Bulletin 4.
13. L. J. Pignataro. Weaving Area Operations Study. NCHRP Project 3-15, 1971.

SIMPLIFIED APPROACH TO MODELING FREEWAY OPERATIONS AND CONTROL

Brian L. Allen, Department of Civil Engineering, McMaster University; and
C. Jen Liew, De Leuw, Cather and Company of Canada

This paper develops a simplified model for simulating freeway operations influenced by entrance-ramp metering or closure. The model's application to a real freeway corridor is demonstrated. The model is based on the assumption that the entire corridor can be adequately represented by only 2 routes interconnected by equally spaced entrance and exit ramps. Optimal control is achieved by minimizing total corridor travel time. The effectiveness of 3 control strategies (entrance-ramp metering, entrance-ramp closure, and total interchange closure) is investigated. Traffic flow on a real freeway corridor was simulated with this model. The model compared favorably with observed conditions. When the effects of the 3 control strategies were investigated, freeway flow rates resulting from optimal control conditions were found to be nearly identical for each strategy. Identifying the optimal flow rate permitted accurate calibration of the model and reliable results. The model can be useful for initial planning evaluations. Data requirements for the model are minimal, and its application is straightforward.

• TRAVEL demand continues to increase, and with it, congestion on urban freeways spreads. This spread can be stopped or slowed by exercising some form of restrictive control. One common form is limiting access to freeways by either closing or metering entrance ramps.

Entrance-ramp metering has been widely accepted and successfully implemented in such freeway corridors as the Eisenhower Expressway in Chicago (1), the Gulf Freeway in Houston (2), the Van Wyck Expressway in New York City (3), and several freeways in the Los Angeles area (4). Vast amounts of monetary and human resources have been spent in metering research, development, and implementation (5, 6, 7).

Freeway-entrance-ramp closure has not been so widely accepted although it appears to be gaining in popularity as existing corridors become more congested. Several operating agencies have closed entrance ramps during peak travel periods, and usually they have had successful results (8, 9, 10). Lack of wider application seems to be because of the method's lack of political popularity, misunderstanding of its potential uses and benefits, and an absence of reasonable locations in which to implement it.

Detailed design of the method, and evaluation of its effectiveness for improving traffic operations, have proved to be a time-consuming and difficult task. To alleviate this burden, a significant proportion of development effort has been expended to provide sophisticated analytic models. These models simulate traffic flow on a freeway or in a corridor subject to a specified ramp-control strategy (11, 12, 13). Usually the models require extensive and accurate data input for successful operation, and, not surprisingly, such data are seldom readily available. For example, the FREQ model series (14, 15, 16) requires the user to supply complete details on freeway physical features, origin-destination patterns of traffic, and metering rates for all entrance ramps. By the time one considers, say, 30 different freeway subsections and 12 time intervals during the peak period, the magnitude of information required is formidable. Undeniably, that amount of detail is necessary if one is to place any degree of confidence in the final design of a control strategy. However, use of such techniques for preliminary analyses of freeway control seems impractical. There appears to be a need for a simplified technique that could be applied, for example, when an operating agency wished to ascertain the need for more detailed analyses on existing or future freeways for which comprehensive traffic data did not already exist.

In this paper, an analytic model is proposed that will fill the need dictated by such an application. The model requires a minimum amount of data for operation, gives reliable results, and serves as a useful first approximation of the detailed design of a freeway-control strategy. In addition, it permits direct comparison of the potential effectiveness of 3 control methods:

1. Entrance-ramp metering,
2. Entrance-ramp closure, and
3. Interchange (entrance- and exit-ramp) closure.

We suggest that this model can be applied directly to preliminary control and deficiency studies of existing freeway corridors and to similar studies for freeway corridors that are being planned or designed. Only a simple trip length distribution for the freeway corridor, speed-flow relationships for the freeway and surface streets, and freeway interchange spacing are required as data input. Numerical output can be used to suggest required metering rates, entrance ramps requiring closure, and optimal interchange spacing. The need for and the effectiveness of the 3 control methods can be directly ascertained.

MODEL DEVELOPMENT

A detailed description of model development is available elsewhere (17). With the goal of a simplified model in mind, we chose the freeway corridor representation shown in Figure 1. It consists only of 2 parallel routes, route 1 (freeway) and route 2 (city streets), interconnected by equally spaced access links (interchanges). All trips in the corridor are generated on route 2 and are destined for some point downstream that also is on route 2. They can enter route 1 on the entrance ramps and can exit by using the exit ramps. These entrances to and exits from route 1 may be selectively closed to permit investigation of the effects of entrance-ramp closure and total interchange closure strategies. $m - 1$ is the number of adjacent entrance ramps that will be closed; ms is the spacing between adjacent accessible entrance ramps.

To enable representation of entrance-ramp metering, one must impose a toll, δ ($\delta > 0$), at all accessible entrance ramps. This toll is considered to be in the form of a travel cost (time) penalty for each trip entering route 1. It represents the wait in queue behind a metering signal.

The segments of route 1 and route 2 between 2 adjacent access links are cells. The corridor comprises a series of individual cells, connected at common access links. Trips begin in an origin block containing the corridor segment between 2 adjacent accessible entrance ramps and terminate in a downstream destination block similarly defined. Each block contains $m + 1$ ramps and is ms long. The distance between corresponding ends of the origin and destination blocks is ns.

Within an origin block, x is the distance measured downstream between the first available entrance link and any specified origin within that block. Similarly, y_k is the distance measured upstream between the last available access link in destination cell ($n + k$) or the kth cell in a destination block and any destination within that block where $k = 1, 2, \ldots, m$. X and Y are the respective distances to these origins and destinations measured from some arbitrary point upstream. The trip length, L, therefore is $Y - X$, the distance between the origin and destination of any trip.

The travel cost per unit of distance of travel on route 1, c_1, is an increasing function of the flow on route 1, f_1. The travel cost per unit of distance of travel on route 2, c_2, is assumed to be independent of the flow on route 2. As shown in Figure 2, $c_1 < c_2$ over rates of flow expected under control conditions.

It is assumed that travelers, because they are aware of the costs of using alternate routes, choose paths with the lowest cost. Travelers making short trips would find

Figure 1. Transportation corridor.

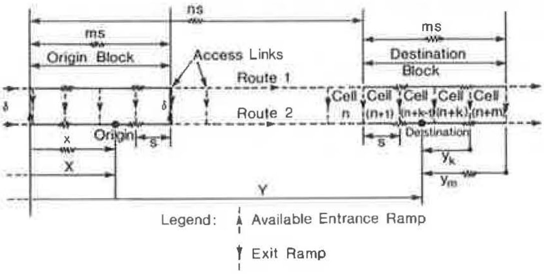


Figure 2. Travel cost per unit of distance of travel on routes 1 and 2.

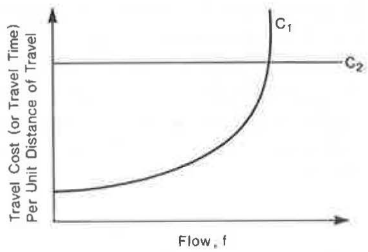


Figure 3. Average distance traveled and total average travel cost for short, intermediate, and long trips.

	Short Trips	Intermediate Trips	Long Trips
	$0 \le L \le ns-\mu\Delta$	$(n+k-1)s-(k-1+\mu)\Delta \le L_k \le (n+k)s-(n+\mu)\Delta$	$(n+m)s-(m+\mu)\Delta < L < \infty$
$f_1(L) =$	0	$\frac{1}{2m}\sum_{k=1}^{m}\{2(n+k-1)(L_k-ns+2\mu\Delta) - (s-2\Delta)(k^2-k) + 2\Delta(k-1)(n-\mu)\}$	$L - \frac{(m+1)}{2}s\frac{c_1}{c_2}$
$C(L) =$	c_2L	$\sum_{k=1}^{m}[c_2L_k - \frac{(\mu+k-1)}{m}(c_2-c_1)(L_k-ns+2\mu\Delta) + \frac{1}{2m}\{\mu(c_2-c_1)\mu\Delta + (k-1+\mu)(c_2-c_1)(k-1+\mu)\Delta + (k-1)(c_2-c_1)[(k-1)(s-2\Delta)+(s-\Delta)]\}]$	$c_1L + \delta + \frac{(m+1)}{2}(c_1+c_2)\Delta$

$$\Delta = \frac{s}{2}(1-\frac{c_1}{c_2}); \quad \mu = n-\frac{s}{s(c_2-c_1)}, \text{ where } 0 \le \mu \le 1$$

Figure 4. Hamilton-Toronto freeway corridor.

Modified Interchanges	Adjusted Spacing (Miles)
1	
2	4.95
3	4.10
4	3.05
5	2.80
6	2.44
7	1.50
8	1.95
9	1.57

that route 2 cost the least time because they would avoid backtracking, queuing at entrance ramps, and extra travel to and from the freeway. Travelers making long trips would find that route 1 cost the least time even with these penalties. Unfortunately, travelers making trips of intermediate length cannot be assigned so easily. Depending on the location of the origin and destination within the blocks, these travelers might use either route 2 or route 1. All trips will be classified as being either short, long, or intermediate.

If all trips can be assigned to the corridor, route flows can be computed. If route flows can be computed, then c_1 and average total travel cost for trips of length L, C(L), can be determined.

To enumerate the number and pattern of trip origins and destinations, we defined trip density function as g(L). There are g(L) trips originating in the corridor segment (X, X + dL), destined for the segment (Y, Y + dL). Thus g(L)dL trips per unit of length are generated at any point along the corridor. By using the average travel cost computed for each of the 3 trip length ranges as shown in Figure 3, integrating $C_i(L)g(L)dL$ over all trip lengths in range i, and summing the 3 numbers, one can calculate total travel cost per unit of corridor length. Similarly, one can compute f_1 by integrating average travel distance, $f_1(L)$, shown in Figure 3, over the 3 trip length ranges.

When total corridor travel cost has been determined, optimization can start. Optimization involves choosing the appropriate metering rate, entrance-ramp closure configuration, or interchange spacing that minimizes cost.

MODEL APPLICATION

It is obvious that each expression in Figure 3 contains 3 unknowns, c_1, δ, and s. Even if both s and δ were fixed and known, c_1 and f_1 would be interrelated. Consequently, an iterative procedure must be used for solution. One must first compute values of f_1 by assuming various values of c_1. The known function $c_1(f_1)$ can be equated with those values and the intercept of the 2 functions will yield the correct c_1. Then all expressions can be solved. To aid in this tedious trial-and-error computation, an interactive computer program was developed, and data from a portion of the Hamilton-Toronto freeway corridor were used as input for a sample computation.

Study Area

The corridor shown in Figure 4 lies between Guelph Line and Highway 427, a distance of about 20 miles (32 km). Route 1 is the 3-lane eastbound portion of the Queen Elizabeth Way (QEW). Route 2 is Highway 2 and all parallel surface streets within 2 miles (3.2 km) of QEW. Perpendicular city streets connect these routes at 14 interchanges. All but the following interchanges have both entrance and exit ramps:

1. Guelph Line, Trafalgar Street, and Mississauga Road, which have additional entrance ramps, and
2. Royal Windsor Drive and Evans Drive, which have no entrance ramps.

The Dixie Road entrance is closer to the Evans Drive exit than it is to the Dixie Road exit. The Highway 427 entrance was outside the chosen study area.

The distance between exit and entrance ramps at any interchange was to be zero to conform with model assumptions. Closely spaced interchanges at Guelph and Walkers Lines, Service and Bronte Roads, Kerr and Trafalgar Streets, and the Evans Drive exit and Highway 427 were combined to form single representative interchanges because they serve a common area and could be considered as single interchanges. The locations of these modified interchanges are shown in Figure 4 by the dashed lines.

The operational characteristics of traffic in this corridor that are partially described by the speed-flow relationship for the QEW shown in Figure 5 were computed from an empirically derived, linear speed-density relationship supplied by the Ontario Ministry

Figure 5. Speed-flow relationship for Queen Elizabeth Way.

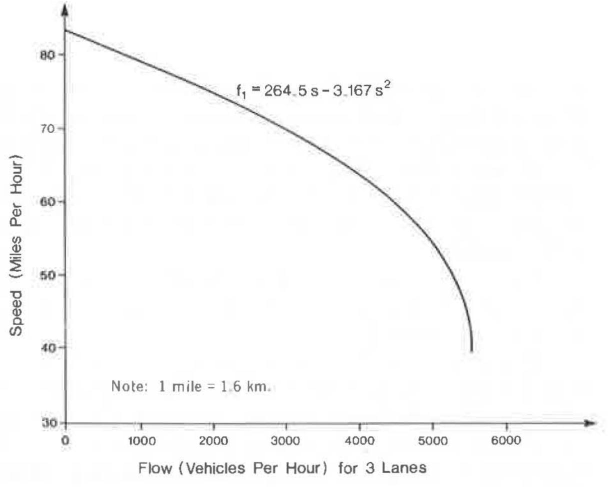


Figure 6. Travel cost-flow relationship for Queen Elizabeth Way.

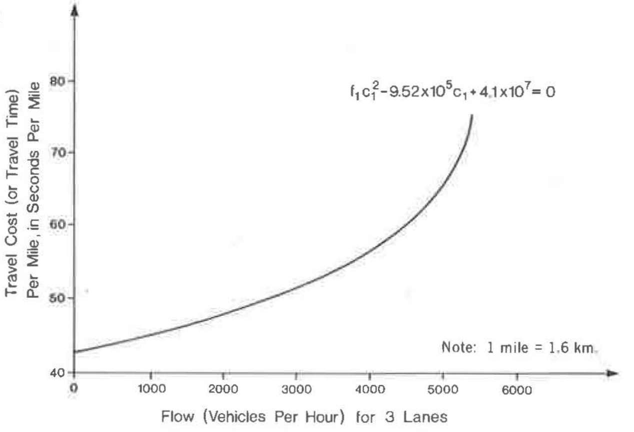
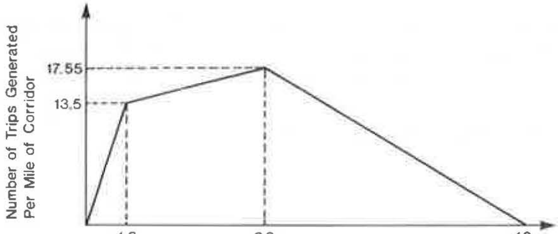
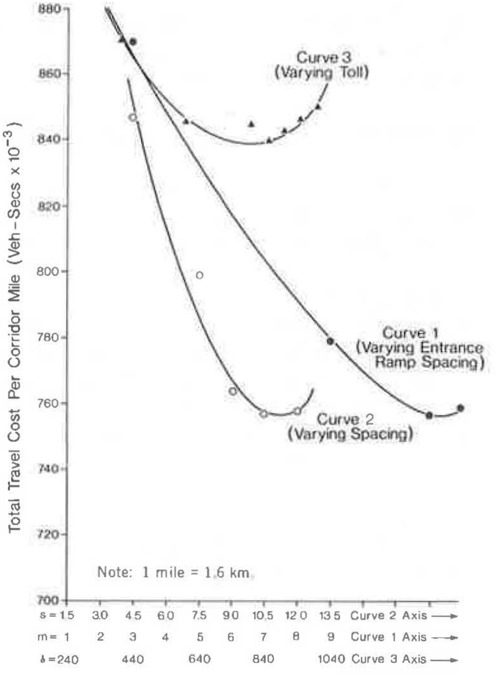


Figure 7. Trip density distribution for Queen Elizabeth Way corridor.



Note: 1 mile = 1.6 km.

Figure 8. Total corridor cost for varying tolls, entrance-ramp spacings, and interchange spacings.



of Transportation and Communications (MTC). Travel time per unit of travel distance on the QEW as a function of flow was obtained from that speed-flow relationship and is shown in Figure 6. Average speed of travel on Highway 2 and parallel surface streets was assumed to be 30 mph (48 km/h). Average speed also was assumed to be essentially independent of flow variations on these streets. These assumptions were considered acceptable because field observations indicated low volume-to-capacity ratios and because any additional flow diverted from the freeway should not affect significantly the speed on these streets.

The magnitude and origin-destination patterns of traffic in this corridor were described by developing the trip density distribution shown in Figure 7. The distribution was computed by using data obtained from a study carried out in the Toronto area (18, p. 53).

People making trips generated in this corridor were assumed to have to travel an additional 2 miles (3.2 km) on lateral surface streets (those connecting routes 1 and 2) if they were assigned route 1 paths. Basing our calculations on an average speed of 30 mph (48 km/h), we assumed that this distance would add a 4-min penalty. Knowledge of existing corridor characteristics indicated that a spacing of 1.5 miles (2.4 km) between interchanges would be most representative of the critical section from Mississauga Road to Highway 427.

Results

All of the information on the chosen corridor was used as input into the computerized model. Results from this application are shown in Figure 8. Curve 1 is the total cost due to varying the spacing of available entrance ramps. It indicates that a spacing of 18 miles (29 km) between available entrance ramps (or closing 12 adjacent entrance ramps) would minimize the total cost of travel to all users in this corridor; the flow on QEW would be 5,100 vehicles per hour (vph).

Curve 2 is the total cost due to varying the spacing of entrance and exit ramps (interchanges). In this case, an interchange spacing of 10.5 miles (16.9 km) would minimize total user cost in this corridor; the flow on QEW would be 5,000 vph. Minimum total cost obtained by varying interchange spacing was not significantly different from the minimum total cost obtained by varying entrance-ramp spacing.

Curve 3 is the total cost obtained by varying the toll imposed on all users entering the QEW. A toll of 900 sec (or an additional penalty of 660 sec) would minimize total user cost in this corridor; the flow on QEW would be 5,100 vph.

Additional runs from this computerized model in which the trip density function and the penalty charged to all users assigned route 1 paths confirmed that a flow of 5,000 to 5,100 vph on the QEW would give the minimum total user cost regardless of the freeway-control strategy used. This flow range corresponds with MTC field observations of optimal travel conditions on the QEW through the critical section.

However, the recommendation for optimum spacing of entrance and exit ramps and entrance-ramp metering rates cannot be realistically applied to this corridor because the chosen 20-mile (32-km) corridor is relatively short. To accommodate this and to make the application more meaningful, trips with lengths greater than 20 miles (32 km) should be considered as external through trips that make up only a constant through flow on the QEW.

After thorough consideration of trip characteristics in this corridor, an external flow of 2,500 vph was computed. A new trip density distribution with a maximum trip length of 20 miles (32 km) was derived from the previous distribution by deleting the portion with trip lengths greater than 20 miles (32 km). The modified input was then fed into the computerized model. The results are as follows (1 mile = 1.6 km):

m	s (miles)	δ (sec per vehicle)	f_1 (vph)
1	1.5	240	5,200
1	3.0	240	5,000
2	1.5	240	5,100

The results indicate that an entrance-ramp spacing of 3 miles (4.8 km) (m = 2, s = 1.5) will reduce the flow to 5,100 vph, whereas an interchange spacing of 3 miles (4.8 km) (m = 1, s = 3.0) will reduce the flow further to 5,000 vph. Both of these flow rates are within the optimal range.

From this second application, one can recommend that some form of freeway-ramp-control strategy be implemented between Erin Mills Parkway and the Dixie Road interchange because of the shorter spacing. If ramp closure is preferred, then the entrance ramps (and exit ramps, if necessary) at Mississauga and Dixie Roads may be closed during the morning peak period to effect the desired optimal spacing.

COMMENTS AND CONCLUSIONS

The simple model of freeway corridor operations and control reported here most certainly will be subject to criticism. The simplifying assumptions used to decompose a complex system of interdependent variables into an extremely simple one are obviously suspect. For example, there never has been a corridor in which all traffic origin-destination patterns were identical along its entire length; neither will there ever be a corridor in which the physical characteristics of the roadways are invariant over length. The formulation of the speed-to-flow or travel-time-to-flow relationships also is open to question. Although no one can strenuously argue that the form used to represent travel on a freeway (route 1) is incorrect, the independence of travel time on flow on city streets is at least a dubious simplification. Oversaturation of critical signalized intersections in the street network could very quickly obviate any benefits realized on the freeway. Finally, the assumption of constant flow along the freeway, regardless of the number of available entrance ramps, is strictly incorrect. If, for example, every second entrance were closed and exits were open, flow would obviously decrease in the subsections immediately downstream of the exit ramps. The equations in Figure 3 that were used to compute travel times on the corridor are also strictly incorrect.

Despite these severe shortcomings, results from the example application seem to indicate, on a gross scale, a strong correlation between actual and simulated conditions. Most importantly, the method reliably predicts flows generated in the most critical sections of the freeway. Field observations also confirm that the model accurately predicts the optimal flow rate for critical freeway sections, that is, the maximum rate of flow that can be maintained without severe travel time increases. These results indicate that the method proposed here can be taken more seriously than we first thought. Because of this, one can also look seriously at several other interesting conclusions drawn from the example.

Perhaps the single most important observation concerns the optimal flow rates computed for critical sections. Results indicated a difference of only 2 percent between the optimum flows computed for each control strategy. This observation not only is intuitively appealing but also has important practical implications. When implementing a given control strategy, one should exercise control so as to obtain the prescribed optimal flow rate on the freeway. Although in practice there may be slight variations in that flow, it appears that sensitivity to minimizing total corridor travel cost would be minimal. Adherence to this procedure would reduce considerably the effort required to provide a final control design.

An application would consist of using the observed (or calculated) trip density function for a wide range of interchange spacings to yield the minimum total corridor travel time. The flow at that minimum would be chosen as the optimal flow. Because a proportion of the total trips on the freeway are likely to be through trips, the optimal spacing inferred by this first computation should be ignored. The trip density function then should be truncated to remove the cumulative influence of those trips and should be replaced as a constant nonadditive flow. The revised density function should then be used in the model to obtain the interchange spacing that yields the optimal flow obtained from the first computation. The spacing thus computed would be the recommended optimal spacing. Using this procedure, one can obtain general recommendations for control by total interchange closure, entrance-ramp closure, or entrance-ramp metering.

Although such a procedure may sound complicated and time-consuming, it is simple and easy to perform with a computer. In addition, the results are extremely easy to interpret. Identification of critical or potentially critical sections simply requires that one compare optimum spacing to existing spacing. If optimum spacing is greater than existing spacing, one should design improvements accordingly.

Data requirements for using this procedure are minimal. Trip length distributions are usually available from operating or planning agencies for almost every major urban corridor, and an indication of the proportion of through trips is obtained easily from a license plate survey or simple truncation of the trip length function. Together with the addition of travel-time functions, these are the only data required to obtain an indication of the degree of control required on the corridor.

Admittedly, this procedure could not be used for detailed design of a control scheme. Although rates established for entrance-ramp metering are unlikely to be equal for all ramps within critical freeway sections, the model results would indicate required rates. The spacings recommended for either accessible entrance ramps or interchanges could not be obtained precisely on a real corridor, but close approximations are usually possible. Finally, no detail concerning operations on the adjacent street network is used in the model so that final consideration of storage lane needs, revised signal timing, and intersection signing could not be established. However, the procedure could be used as a workable first approximation of control requirements. Perhaps it could be used in conjunction with standard deficiency studies, or it could be associated with planning and design procedures in which the availability of detailed data is limited. In any of these cases, final control specifications are not required, so use of more complicated models would not be warranted.

In addition to specifying approximate freeway control needs, output from the model would also be useful for comparing the relative effects of the 3 control strategies. Although such comparisons would be qualitative, they would be useful when one is considering trade-offs between strategies or contemplating a combination of control modes. Current activities should be expanded and continued so that a better understanding of various control modes will result.

We suggest that the model reported here offers considerable advantages over currently available methods for examining freeway corridor operations and control. Although it is not comprehensive in nature, it provides reliable indications of the extent and degree of control required and demands very little in the way of data preparation and output interpretation. It provides an essential link between awareness of problems, understanding the applications of various control modes, and final implementation of control.

ACKNOWLEDGMENTS

We thank J. Thomas and R. Walshaw, formerly of the Ontario Ministry of Transportation and Communications, for their assistance in providing data used in the example applications.

This work was made possible by sponsorship of the National Research Council of Canada.

REFERENCES

1. Chicago Area Expressway Surveillance and Control. Illinois Division of Highways, July 1971.
2. M. E. Goolsby, E. Merrell, and W. R. McCasland. Freeway Operations on the Gulf Freeway Ramp Control System. Texas Transportation Institute, Texas A&M Univ., Rept. 24-25, 1969.
3. L. Pignataro and K. Crowley. The Van Wyck Expressway Experiment. Traffic Engineering, Vol. 41, No. 4, Jan. 1971.
4. Los Angeles Area Freeway Surveillance and Control Project. California Division of Highways, 3rd Annual Report, Feb. 1973.
5. L. Newman, A. Dunnet, and G. Meis. Freeway Ramp Control—What It Can and Cannot Do. Traffic Engineering, Vol. 39, No. 9, June 1969.
6. W. Wainwright and R. Jain. Techniques of Ramp Control. Univ. of Connecticut, C.E. 72-47, Feb. 1972.
7. P. Everall. Urban Freeway Surveillance and Control: The State of the Art. Federal Highway Administration, Nov. 1972.
8. R. Loutzenheiser. The Effects of Ramp Closure on Freeway Operation During Morning Peak Periods. Texas Transportation Institute, Texas A&M Univ., Rept. 139-9, Aug. 1971.
9. L. Newman, A. M. Dunnet, and G. J. Meis. An Evaluation of Ramp Control on the Harbor Freeway in Los Angeles. Highway Research Record 303, 1970, pp. 44-55.
10. E. Gervais and W. Roth. An Evaluation of Freeway Ramp Closure. Michigan State Highway Department, April 1965.
11. H. Payne. Aggregate Variables Models of Freeway Traffic. Univ. of Southern California, Aug. 1971.
12. Y. Makigami et al. The Freeway Model. In Bay Area Freeway Operations Study—Final Report, Part 1, Univ. of California, Berkeley, Aug. 1970.
13. M. Orthlieb et al. Progress Toward a Freeway Corridor Model. In Freeway Operations Study—Phase III, Univ. of California, Berkeley, June 1973.
14. W. Stock et al. The FREQ3 Freeway Model. In Freeway Operations Study—Phase III, Univ. of California, Berkeley, June 1973.
15. M. Eldor and A. May. Cost-Effectiveness Evaluation of Freeway Design Alternatives. In Freeway Operations Study—Phase III, Univ. of California, Berkeley, June 1973.
16. J. Wang and A. May. Analytic Techniques for Ramp Control. In Freeway Operations Study—Phase III, Univ. of California, Berkeley, Interim Rept. 2, June 1972.
17. C. J. Liew. Entrance Ramp Closure as a Means for Controlling Traffic in a Transportation Corridor: Theory and Application. McMaster Univ., master's thesis, May 1974.
18. Metropolitan Toronto and Region Transportation Study. April 1966.

MATHEMATICAL EVALUATION OF TRAFFIC CONTROL ALTERNATIVES FOR RESTRICTED FACILITIES

E. A. Gonzalez,* De Leuw, Cather and Company; and
R. C. Loutzenheiser and E. C. Carter, University of Maryland

As part of a continuing research study directed to alleviate traffic congestion in the Baltimore Harbor Tunnel, the Howard policy-iteration technique was applied off-line to different traffic control alternatives. Three data-acquisition stations were used inside the tunnel for control purposes. One was at the tunnel's bottleneck, and 2 were upstream of this location. The value of the traffic concentration, which was used as the control variable, at these 3 stations defined 1 out of 18 possible states of tunnel traffic flow. State transitions in the system were assumed to occur every 30 sec. Extensive data collected in the Baltimore Harbor Tunnel were used to determine the state transition probabilities of the system under each alternative. The rewards associated with state transition were obtained by applying a model that considers actual flow at the bottleneck and average speed associated with the flow during each transition interval. Five different control alternatives were considered. One of the alternatives was no control; the remaining 4 were 2- to 4-min cycle lengths of a traffic signal located upstream of the tunnel entrance.

•NORMAL operation of urban freeways is frequently affected by excessive traffic demand. Most drivers have experienced overcrowded highways and delays during morning and evening peak periods. The limited capacity of a highway network is often exceeded by the number of vehicles trying to use the roadway during these periods. As a result, congestion develops and is accompanied by stop-and-go driving conditions. These conditions, in turn, permit fewer vehicles to be served in a given time period. Congestion is significantly more severe for restricted facilities such as tunnels.

Congestion is a daily routine for the Baltimore Harbor Tunnel Thruway (Fig. 1). Traffic is frequently backed up for 2 miles (3.2 km). The problem is especially crucial on weekends when backups extend for more than 4 miles (6.4 km) and affect Interstate highways. This congestion is aggravated further by the fact that the thruway is a toll facility where exit is completely restricted until the toll plaza has been reached.

Research efforts have established that traffic control is an appropriate means to not only improve traffic-flow characteristics during congested periods but also prevent traffic from reaching states of potential congestion (1).

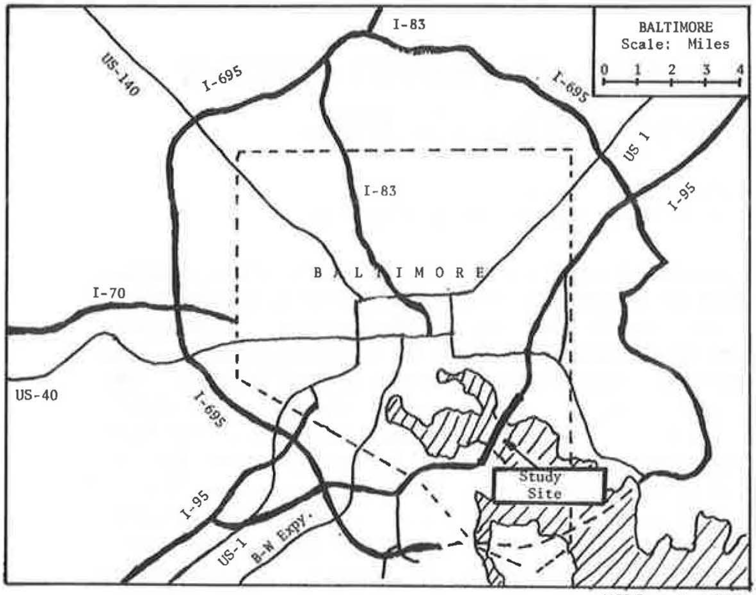
This paper discusses the feasibility of a stochastic or probabilistic approach to reducing traffic congestion on restricted facilities. This probabilistic approach uses the Howard policy-iteration method (2, 3, 4), which is based on the Markov process with rewards. The method was applied to traffic-flow data collected in the Baltimore Harbor Tunnel (5, 6).

LITERATURE

Because we analyzed traffic control alternatives by a Markovian approach in this re-

*Mr. Gonzalez was with the University of Maryland when this research was performed.

Figure 1. Baltimore Harbor Tunnel.



NOTE: 1 mi = 1.6 km.

search, we dealt with Markovian analysis and control of traffic congestion in a literature review. Because the need for control largely is due to traffic congestion, we undertook a short examination of the causes and consequences of congestion (5, 6).

When the relationship between variables is probabilistic or random, stochastic models are used. These models can be independent if the outcomes of the experiments do not influence each other and they can be Markovian if the outcome of 1 experiment is directly dependent on the preceding experiment.

Markov processes have been used to correlate successive headways in traffic streams with traveling platoons (7) and in traffic-merge problems (8). In 1967 Jewell (9) recognized the potential of Markovian approaches to traffic-flow theory. In 1972 Haefner and Warner (10) applied the Howard policy-iteration method with rewards to a hypothetical traffic control case. In 1973 Carter and Palaniswamy (5) formulated a conceptual approach to the analysis of traffic control alternatives in the Baltimore Harbor Tunnel. This formulation was further explored the same year by Palaniswamy (11) who suggested a reward structure that could be applied when data became available.

During the studies described by Carter and Palaniswamy (5) and Palaniswamy (11), the Baltimore Harbor Tunnel was divided into 4 major sections:

1. Queue area (upstream of and including the toll plaza),
2. Merge area (immediately downstream of the toll plaza),
3. Ramp area (joins the merge area to the tunnel), and
4. Tunnel (restricted facility).

The tunnel (Fig. 2) was further divided into downgrade, level, and upgrade sections.

During heavy demand, the critical bottleneck was at the foot of the upgrade; this finding agrees with the results obtained in previous New York tunnel studies (12, 13, 14, 15, 16, 17, 18). The rest of the study therefore was directed toward improving traffic flow at the bottleneck.

Data for earlier studies at the Baltimore Harbor Tunnel were collected at 7 stations (Fig. 2). Station 5 was located at the critical bottleneck. A description of the data collection equipment, the detail setup, and problems encountered is given elsewhere (5).

Figure 2. Plan and profile views of Baltimore Harbor Tunnel with station locations.

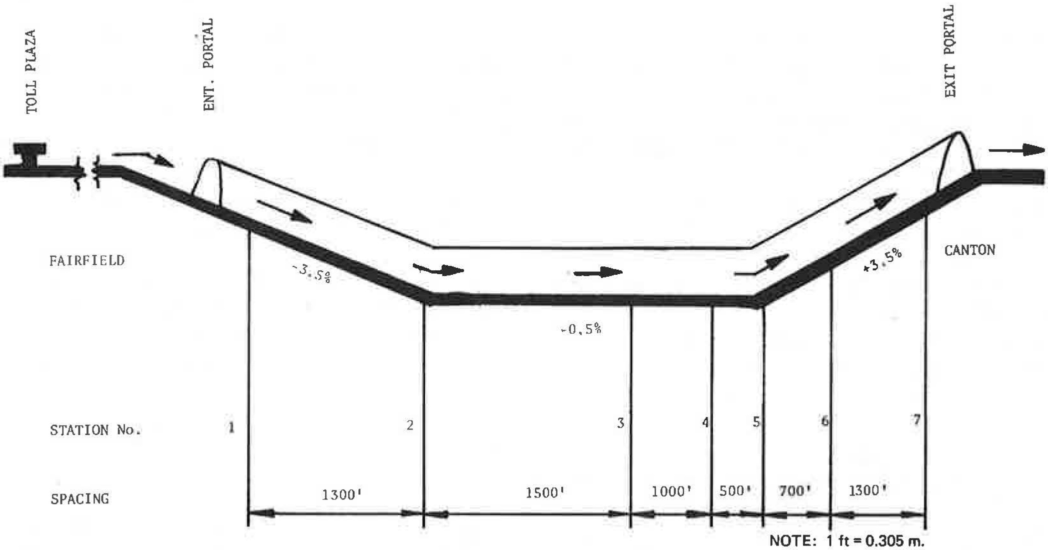


Figure 3. Location of signals and signs used for alternatives.

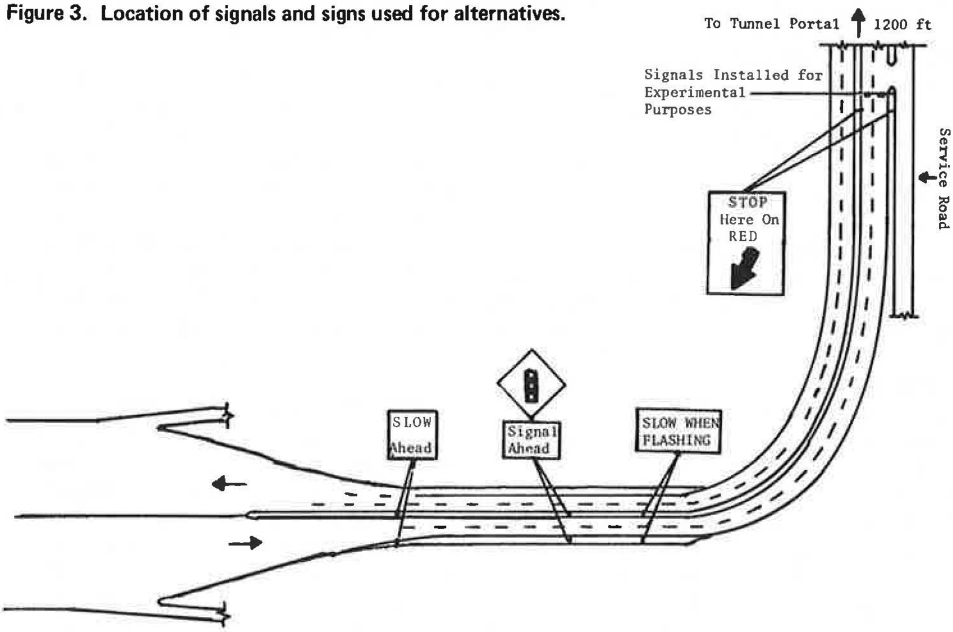


Table 1. Description of alternatives.

Alternative	Cycle Length (sec)	Green (sec)	Amber (sec)	Red (sec)
1	No control		0	0
2	120	109.2	3.6	7.2
3	160	147.2	4.8	8.0
4	180	169.2	3.6	7.2
5	240	225.6	4.8	9.6

STUDY METHODOLOGY

Gonzalez (6) gives a brief description of the Howard policy-iteration method with rewards. A more complete treatment of the subject can be found elsewhere (2, 3, 4, 19).

It is convenient to think of a Markov process as a sequence of states through which a system passes stochastically at successive points in time (8). The states are the various possible conditions in which the system might be at any instant of time. Each state must be uniquely described by the values of a variable or set of variables. When the values of the describing variables change from those of one state to those of another, a state transition is said to have occurred. State transitions can be considered to occur at discrete time intervals.

As the system passes from state i to state j, it earns a reward the value of which depends on states i and j. If several alternatives are examined, each alternative will have its own state transition behavior and its own reward. The Howard policy-iteration method finds the best alternative associated with each state of the system, given the transition probabilities and the rewards associated with each state.

Description of Alternatives and Data Collection

Metering was accomplished with a pretimed signal located in the ramp about 1,200 ft (366 m) upstream of the tunnel entrance. Figure 3 shows a general layout of the site and the warning signs used in conjunction with the metering experiments. The 5 alternatives used are given in Table 1.

Because of the experiences of the Port Authority of New York and New Jersey with the 1-min cycle and because of the high capacity of the Baltimore Harbor Tunnel, it was felt that the 2-min cycle should be the minimum.

Data obtained for each of the alternatives were collected under the same circumstances of heavy traffic demand. Demand was determined by the length of the queue upstream of the tunnel proper. The queue had to extend to the point where the SLOW AHEAD sign was located. If this condition of demand was not met, no data were collected. Because the data collection period coincided with the energy crisis, the condition of not enough demand was the rule rather than the exception.

When the Howard policy-iteration technique is used, the system must dwell in as many states as possible so that state transitions can occur over an ample range. This was insured in this study by the way the traffic metering was started every day of data collection. Initially, each alternative was carried out and data were collected without discontinuity in or stoppage of traffic before metering began. This resulted in congested starting conditions. A medium level of starting congestion was obtained by stopping the traffic at the signal for 90 sec. This stoppage greatly relieved the state of congestion inside the tunnel, but the time was not long enough to have the tunnel completely cleared of vehicles. The lowest level of congestion was obtained by stopping the traffic for as long as was necessary to allow the traffic already in the tunnel to clear station 5, the bottleneck location. An observer located at this station would radio to a police car adjacent to the signal when this occurred, and traffic then would be released. This procedure was followed for each of the alternatives to provide, when possible, similar conditions for each alternative. When similar conditions were attained, data were collected. A large amount of data had to be collected to obtain the transition probabilities associated with each alternative. If a small sample was used, possible state transitions might not be observed; in the final analysis these transitions would be treated as nonexistent.

Where measurements are to be made is another important factor. It is recommended that 1 of the locations be at the bottleneck because capacity is lowest at this point and shock waves that lead into congestion most likely will originate there.

Description of the system, state, and control operation becomes better as the number of data collection stations increases. In the Baltimore Harbor Tunnel 3 stations were used: 1 in the downgrade (station 1), 1 in the level section (station 3), and 1 at the bottleneck (station 5). The equipment used in the data collection consisted of high-intensity light sources placed on the upper portion of the side wall of the tunnel and directed at photoconductive cells under the pavement. The data were recorded and stored on magnetic tape. All of the details concerning collection, storage, and manipulation of data, including several problems encountered in the installation of the data acquisition system, are explained by Carter and Palaniswamy (5).

State Definition

Because concentration, K, is a quantitative measure of congestion (20), it is appropriate to use it as the control variable. States then can be defined in terms of concentration values at certain locations within the tunnel.

If a large number of states are used, a complicated and costly control algorithm could result (10). On the other hand, if few states are used, the description of the tunnel's state of congestion can be obscured to the point where situations requiring control would be overlooked. Such a case would be a nonoptimal situation.

A careful study of the volume-concentration-speed (Q-K-V) relationships for the traffic stream in the Baltimore Harbor Tunnel revealed that concentrations of 55 vehicles/mile (34 vehicles/km) and more were typical of unstable conditions; concentrations of 40 vehicles/mile (25 vehicles/km) were characteristic of stable, uncongested flows. The final state definition was obtained by combining 3 substates, 1 from each of the 3 stations.

For station 1, 3 possible substates were defined (1 vehicle/mile = 0.62 vehicle/km):

1. $0 < K < 40$ vehicles/mile,
2. $40 \leq K \leq 55$ vehicles/mile, and
3. $55 < K$ vehicles/mile.

For station 3, 2 possible substates were defined (1 vehicle/mile = 0.62 vehicle/km):

1. $0 < K < 55$ vehicles/mile, and
2. $55 \leq K$ vehicles/mile.

Station 5 is critical to the operation of the whole system and therefore was assigned 3 possible substates (1 vehicle/mile = 0.62 vehicle/km):

1. $0 < K < 40$ vehicles/mile,
2. $40 \leq K \leq 60$ vehicles/mile, and
3. $60 < K$ vehicles/mile.

The higher limit for station 5 reflects the fact that observed concentrations at station 5 were consistently higher than they were at the other stations.

According to this scheme, the number of states for the tunnel as a whole is 18 ($3 \times 2 \times 3 = 18$). Figure 4 shows the possible combinations of substates and states. Note that state 1 has the lowest concentration values throughout the tunnel, and therefore reflects the least congestion. State 18 reflects the most congestion. This enumeration of states does not necessarily mean that the higher the state number is, the greater is the degree of congestion. For example, it is not necessarily true that state 7 is more congested than state 6. State numbers, then, are more matters of mathematical convenience than they are matters of actual desirability.

Transition Probabilities

When the alternatives were being carried out, all traffic incidents were noted. The time of occurrence and duration of each incident also were recorded. A close comparison of these notes to the time and characteristics of the data stored on magnetic tapes was used to eliminate data that were not directly a result of the specific alternative being tried. This procedure was carried out for each station for each day of data collection. The usable data, in 30-sec averages, were then stored by alternatives on different magnetic tapes. A check was made to verify that the same number of observations was taken at each of the 3 stations and that the observations were taken simultaneously.

Finally, the data were processed to determine the state of the system at any given time interval, $t = T$, and at the next interval $t = T + 30$ sec. From this determination the sample size of each individual state transition was obtained. The state transition probabilities then were calculated by dividing the sample size of the individual state transitions by the total number of transitions from that state.

The rewards associated with these transitions were obtained simultaneously.

Reward Structure

It has been mentioned that the system can be described by 18 states. Each of these states has control alternatives associated with it. When 1 of the alternatives is chosen for a given state i, a decision has been made for that state. The set of decisions for all states is called a policy (4). The optimal policy is that which maximizes the gain, g, or average return, per transition. The object of the Howard process is to define such a policy.

A reward is associated with the transition from one state to another every 30 sec. This reward can be considered to be vehicles processed by the facility, savings in travel time, increased speed or safety, or any other meaningful traffic-related variable or combination of variables.

Drew and Keese (21) suggested a measure of performance that simultaneously involves flow, or volume, and speed. This parameter is called the kinetic energy, E_k, of the traffic stream and is given by the product of flow and speed. The reward structure to be proposed for the Howard process involves flow, speed, and concentration. A further discussion of some traffic-flow concepts will help the reader to understand this structure.

Consider the Q-K and V-K curves for a given facility (Figs. 5 and 6). Maximum flow, Q_M, is reached at a certain value of concentration K_M, optimum concentration, (Fig. 5); K_M at the same time determines the theoretical speed at which the flow is maximized, V_M (Fig. 6). When E_k is maximized the corresponding flow will be Q_M', which is less than Q_M. The concentration K_M' will be less than K_M and the velocity V_M' will be higher than V_M (21).

Regression analysis was used to test the fit of data from the Baltimore Harbor Tunnel to the Greenshields (22), Greenberg and Daow (18), and Underwood (23) V-K models. The Greenshields model was selected on the basis of its estimation of Q_M, V_M, and K_M parameters and according to its coefficient of determination, R^2, and standard error of estimate, S_e. The equation obtained is as follows:

$$V = V_f(1 - K/K_j)$$

where

V_f = free-flow speed [56.8 mph (90.9 km/h)];
K_j = jam concentration [112.8 vehicles/mile (70.94 vehicles/km)];
R^2 = 0.86; and
S_e = 5.79 mph (9.3 km/h).

Some of the special values associated with this equation are

1. Q_M = 1,604 vehicles per hour (vph);
2. V_M = 28.4 mph (45.4 km/h);

Figure 4. Definitions of states.

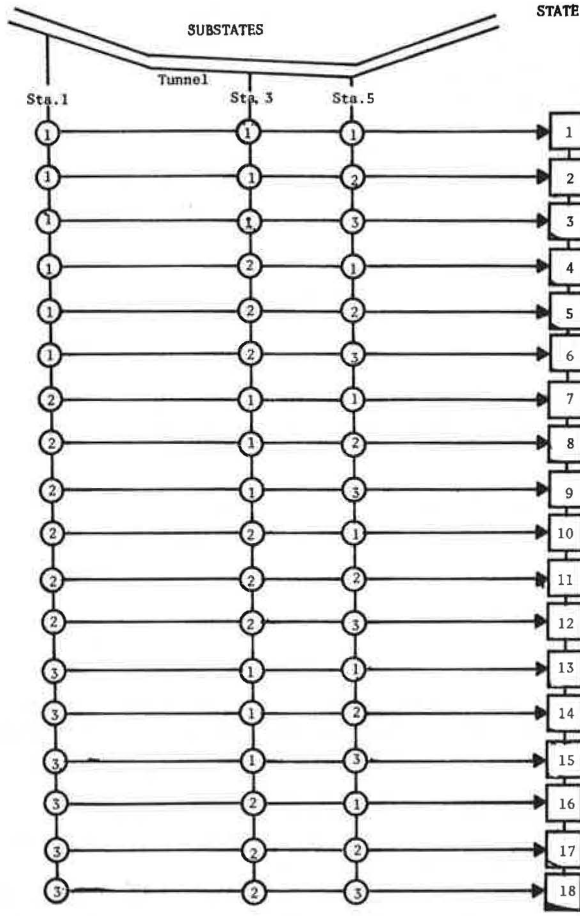


Figure 5. Volume-concentration curve.

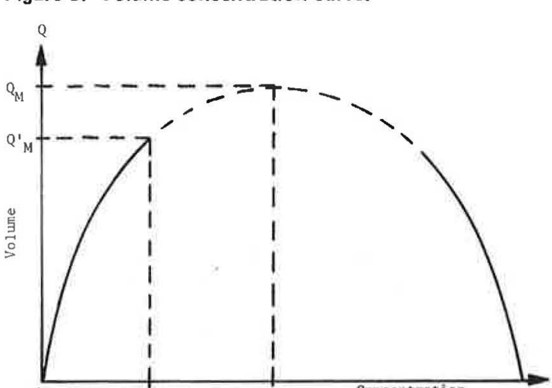
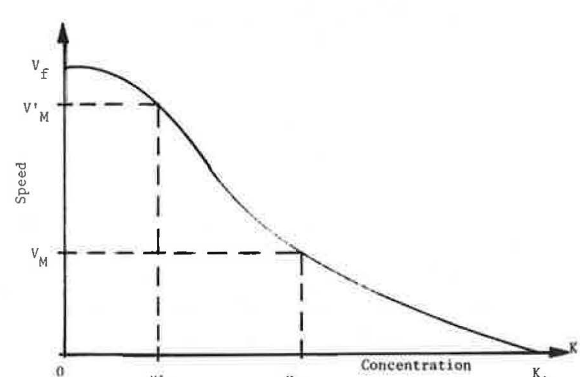


Figure 6. Speed-concentration curve.



3. K_M = 56.5 vehicles/mile (35.03 vehicles/km);
4. Q_M' = 1,425 vph;
5. V_M' = 37.9 mph (60.6 km/h); and
6. K_M' = 37.7 vehicles/mile (23.37 vehicles/km).

The reward, r_{ij}, associated with a transition of the system from state i to state j, is the sum of 2 values:

$$r_{ij} = Q + B \tag{1}$$

where

Q = actual flow processed by the bottleneck during the transition time interval; and
B = bonus, which is dependent on solely the average speed of Q during the same interval.

Maximum bonus, B_M, is assigned to a speed of V_M', the theoretical speed associated with maximum E_k. Its value is given by the following equation:

$$B_M = 2(Q_M - Q_M') \tag{2}$$

Any other average V for the traffic stream in a given interval will have a B that always is smaller than B_M. This B is equal to zero under 2 different circumstances: $V = V_M$ and $V = V_f$.

1. If $V = V_M$, the flow is unstable, and a small increase in traffic demand might be accompanied by a large decrease in speed and a large increase in concentration (11). Therefore, congestion is very likely to occur.
2. If $V = V_f$, extremely high headways between vehicles are more likely to occur, which in turn mean smaller actual flows. It should be noted that V possibly could be greater than V_f. In such cases B = 0.

For any other V values, B is expressed by the following linear relationships:

$$B = \frac{B_M(V_f - V)}{V_f - V_M'}, \quad V > V_M' \tag{3}$$

$$= \frac{B_M(V - V_M)}{V_M' - V_M}, \quad V \le V_M' \tag{4}$$

Note that Eq. 4 implies negative Bs for speeds below V_M. This is logical because slow speeds at the bottleneck are highly undesirable and therefore should be penalized.

For the Baltimore Harbor Tunnel, the special values associated with the equation from Greenshields' model should be used in Eqs. 2, 3, and 4. The maximum bonus was obtained by using Eq. 2.

$$B_M = 2\ (1{,}604 - 1{,}425) = 358 \text{ vph}$$

$$= (\text{about 3 vehicles/30 sec})$$

The relations among Q, K, V, and B for the Baltimore Harbor Tunnel are shown in Figure 7.

Rewards are obtained by adding the observed flow at station 5 and the bonus, which is obtained from Eqs. 3 and 4.

Results and Interpretation

The state transition probability and reward matrices for each alternative are the necessary inputs to the Howard policy-iteration algorithm. Gonzalez (6) developed a computer program to analyze these data. By using the policy-iteration procedure, he was able to obtain the optimal decision matrix given in Table 2. The expected immediate rewards and relative values also were obtained and are given in Table 2.

The elements of the decision matrix correspond to the number of the alternative in the i th state that, in the long run, will maximize g. For example, when the state of the system is 1, the optimal decision is alternative 2, or the 120-sec cycle. Any other alternative will have a reward, but, in the long run, the 120-sec cycle will yield the maximum reward. The same statements hold true for the other states and the associated optimal decision. Note that the optimal policy is made up of 4 different alternatives. Alternative 5, the 240-sec cycle, does not appear.

Gain is g = 14.10 units/30 sec, or g = 14.10 × 120 = 1,692 units/hour. Note that values are in units/hour rather than vph. This is because the value 1,692 does not mean that a flow of 1,692 vph can be expected. g, as is the reward r^m_{ij}, is made up of 2 parts: actual flow and a bonus according to the speeds at the bottleneck.

The 1,692 units/hour reflect a range of conditions that include a volume of 1,692 vph served by the bottleneck at a speed of 28.4 mph (45.7 km/h) in which the bonus equals zero (Fig. 7) or a reduced volume (1,692 - B_M = 1,692 - 360 = 1,332 vph) served at a speed of 37.9 mph (61.0 km/h). Combinations of values between these limits also can occur. Theoretical Q for any given speed will be

$$Q = 1{,}692 - B \tag{5}$$

B is calculated by either Eq. 3 or Eq. 4. For example, at an average speed of 30 mph (48 km/h), the B associated with the traffic stream is given by Eq. 4 as follows:

$$B = 360 \times (30 - 28.4)/(37.9 - 28.4) = 61 \text{ vph}$$

Taking this value to Eq. 5 yields

$$Q = 1{,}692 - 61 = 1{,}631 \text{ vph}$$

Therefore, at 30 mph (48 km/h) a theoretical Q of 1,631 vph can be expected.

It is important to note that, theoretically, volumes even larger than 1,692 vph could be served, but they would be served at speeds lower than those associated with maximum flow [V_M = 28.4 mph (45.7 km/h)]. This is due to the nature of the bonus structure, which assigns negative bonuses to these speeds. For example at a speed of 25 mph (40 km/h), Eq. 3 indicates

$$B = 360 \times (25 - 28.4)/(37.9 - 28.4) = -128 \text{ vph}$$

Figure 7. Relationship among volume, concentration, speed, and bonus values.

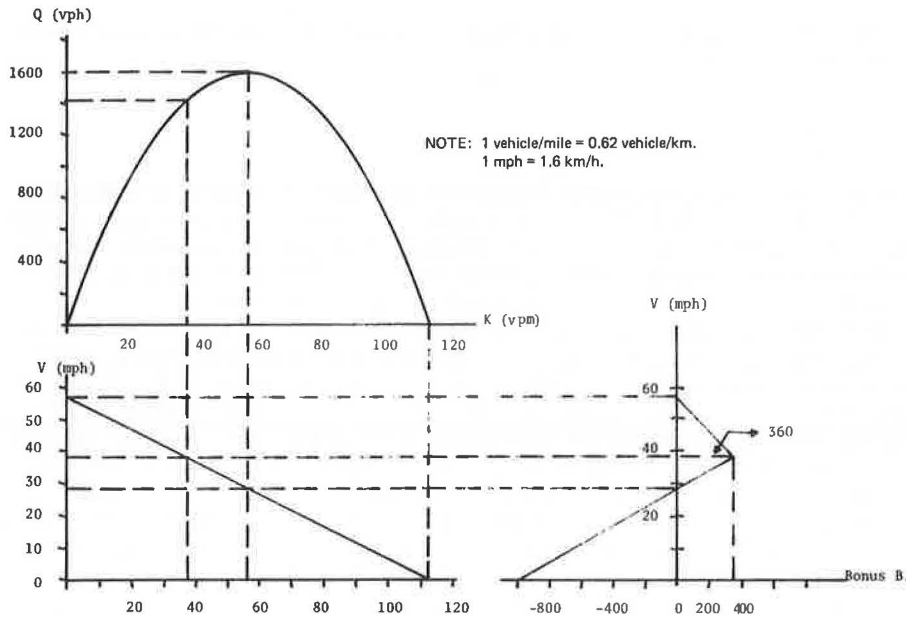


Table 2. Optimal decisions and associated values.

State	Optimal Decision	Expected Immediate Rewards (units/30 sec)	Final Relative Values (units/30 sec)
1	2	14.98	V(1) = 45.45
2	2	14.99	V(2) = 43.86
3	2	12.48	V(3) = 30.30
4	4	14.23	V(4) = 44.38
5	2	12.74	V(5) = 13.32
6	4	10.08	V(6) = 3.91
7	2	14.73	V(7) = 44.25
8	2	15.00	V(8) = 43.71
9	3	11.59	V(9) = 22.37
10	1	13.51	V(10) = 44.22
11	4	12.81	V(11) = 13.17
12	2	10.33	V(12) = 3.10
13	4	16.86	V(13) = 44.81
14	3	13.71	V(14) = 35.33
15	3	9.51	V(15) = 13.53
16	4	11.67	V(16) = 2.18
17	4	10.56	V(17) = 0.68
18	2	9.24	V(18) = 0.00

And the theoretical flow at this speed would be

$$Q = 1{,}692 - (-128) = 1{,}820 \text{ vph}$$

The problem with this flow is that, because it occurs at higher values of concentration at the bottleneck (Fig. 7), vehicles will be packed more closely, and a single slow vehicle in the group might create a general slowdown in the following traffic. That is, a single vehicle might have a shock-wave effect, which in turn might result in a breakdown of flow into stop-and-go conditions. When this situation arises, flow will sharply decrease, and the theoretical expected flow will not be obtained continuously, although it might occur for a certain period of time.

At higher speeds, Figure 7 shows that the concentration is smaller. Headways are larger and the shock-wave effect is more likely to be absorbed. Conditions existing before a vehicle's slowdown are more likely to be restored, and the expected theoretical flow can be obtained. As a result, a smoother flow throughout the facility is more likely to occur.

This is one of the main reasons why the reward structure was chosen in a way that would involve not only actual flow but also traffic speed at the bottleneck. The theoretical maximization of traffic throughput alone might result in a smaller actual flow because of the higher probability of a state of total congestion. This would impede the smooth and continuous movement of traffic through the facility.

It is important to note in Table 2 that gain was obtained with an optimal policy that involved 4 different alternatives that were used according to the state of the system. This gain, therefore, requires a system capable of determining the state of the tunnel at a given moment and transmitting a command to implement the corresponding optimal alternative. This could be achieved by a real-time control system.

CONCLUSIONS

This research has examined the feasibility of applying the Howard policy-iteration method to the evaluation of traffic control alternatives. The system is categorized into 18 different states, which include all of the possible situations encountered by the physical system. Concentration is used as the control variable. The rewards associated with the state transitions are defined in terms of actual flow and average speeds at the bottleneck for 30-sec periods. The Howard method was applied to data collected at the Baltimore Harbor Tunnel for 5 different pretimed metering alternatives (including no control). Because of the lack of on-line hardware, the method could not be tested in real time.

The results obtained for this facility show that

1. The Howard policy-iteration method can be successfully used in evaluating different traffic control alternatives.
2. The optimal policy obtained by this method generally is composed of different alternatives. The results obtained in this research seem to indicate that the policy-iteration method is suitable for application to systems that can continuously monitor the state of the system and subsequently implement the optimal alternative associated with this state.
3. Elimination of nondesirable alternatives is accomplished by the policy-iteration method. These alternatives will not appear in the optimal decision vector. In the case considered in this paper, for example, the 240-sec cycle alternative was eliminated.
4. Normal operation should be used as an alternative. For some states, this situation might be the optimal control alternative.
5. Maximization of expected immediate rewards is not necessarily the best policy in the long run. The long-run criterion should be optimized if the system is to operate for a large number of intervals. This is what the policy-iteration method achieves

when it maximizes the gain.

6. Applying the results obtained by this method to on-line control is the only way to determine with certainty the applicability of the procedure. The theoretical benefits certainly indicate that this should be done.

ACKNOWLEDGMENT

The opinions, findings, and conclusions expressed in this report are those of the authors and are not necessarily those of the Maryland Highway Administration or the Federal Highway Administration.

REFERENCES

1. P. Athol and A. Bullen. Multiple Ramp Control for a Freeway Bottleneck. Highway Research Record 456, 1973, pp. 50-54.
2. R. A. Howard. Markov Models. In Dynamic Probabilistic Systems, Vol. 1, John Wiley and Sons, Inc., New York, 1971.
3. R. A. Howard. Semi-Markov Decision Process. In Dynamic Probabilistic Systems, Vol. 2, John Wiley and Sons, Inc., New York, 1971.
4. R. A. Howard. Dynamic Programming and Markov Processes. M.I.T. Press, Cambridge, Mass., 1960.
5. E. C. Carter and S. P. Palaniswamy. Study of Traffic Flow on a Restricted Facility. Univ. of Maryland at College Park, Interim Rept., Phase 1, 1973.
6. E. A. Gonzalez. A Markovian Analysis of Traffic Control Alternatives for Restricted Facilities. Univ. of Maryland at College Park, MS thesis, 1974.
7. Weiss and Maradudin. Markov Renewal Models in Traffic Flow Theory. Operations Research Center, Univ. of California, Berkeley, 1960.
8. D. R. Drew. Traffic Flow Theory and Control. McGraw-Hill, Book Company, New York, 1968.
9. W. S. Jewell. A Comment on Markov Renewal Models in Traffic Flow Theory. In Vehicular Traffic Science (L. C. Edie et al., eds.), American Elsevier Publishing Company, New York, 1967, pp. 95-96.
10. L. E. Haefner and J. A. Warner. Toward a Markovian Traffic Control Evaluation System. Highway Research Record 456, 1973, pp. 20-27.
11. S. P. Palaniswamy. Optimization of Traffic Flow on a Restricted Facility: A Conceptual Framework. Univ. of Maryland at College Park, PhD dissertation, 1973.
12. K. Crowley and I. Greenberg. Multiple Point Experimentation and Analysis in Single Lane Traffic. Port of New York Authority, Rept. TBR 9-64, 1964.
13. K. Crowley and I. Greenberg. Holland Tunnel Study Aids Efficient Increase of Tube's Use. Traffic Engineering, March 1965.
14. L. C. Edie and R. S. Foote. Effect of Shock Waves on Tunnel Traffic Flow. HRB Proc. 39, 1960, pp. 492-505.
15. L. C. Edie and R. S. Foote. Traffic Flow in Tunnels. HRB Proc. 37, 1958, pp. 334-344.
16. D. C. Gazis. Traffic Flow and Control: Theory and Applications. American Scientist, Vol. 60, July-Aug. 1972, pp. 414-424.
17. D. C. Gazis and R. S. Foote. Surveillance and Control of Tunnel Traffic by On-Line Digital Computer. Transportation Science, Aug. 1969, pp. 255-275.
18. H. Greenberg and A. Daow. The Control of Traffic Flow to Increase the Flow. Operations Research, Vol. 8, 1960.
19. J. G. Kemeney and J. L. Snell. Finite Markov Chains. D. Van Nostrand Company, New York, 1960.
20. J. J. Haynes. Some Considerations of Vehicular Density on Urban Freeways. Highway Research Record 99, 1965, pp. 59-80.
21. D. R. Drew and C. I. Keese. Freeway Level of Service as Influenced by Volume

and Capacity Characteristics. Highway Research Record 99, 1965, pp. 1-47.
22. B. D. Greenshields. A Study of Traffic Capacity. HRB Proc. 14, 1934, pp. 448-477.
23. R. T. Underwood. Speed, Volume and Density Relationships: Quality and Theory of Traffic Flow. Yale Bureau of Highway Traffic, 1961, pp. 141-188.

DESIGN TECHNIQUE FOR PRIORITY-ACCESS RAMP METERING

Saul D. Miller and Harold J. Payne, Technology Service Corporation

One proposed method to provide preferential treatment for buses and car pools on freeways would allow buses and car pools immediate access and would meter the rest of the entering traffic. This paper presents a ramp-metering design technique that is appropriate for a situation in which all ramps under consideration have a priority-vehicle access lane. The technique is optimal in that it minimizes total passenger travel time in a limited corridor. An example design is presented to illustrate the procedure. In addition, conclusions are drawn relating to ramp queues as incentives to make travelers shift to car pools or buses.

•IN THE PAST few years, transportation planners have increasingly emphasized the consideration of alternative transportation modes on freeways. Among these alternatives are various forms of preferential treatment for buses and car pools (2, 3, 5, 6, 9, 10).

This paper is concerned particularly with the mode in which priority vehicles (buses and car pools) are provided preferential access to the freeway and nonpriority vehicles are subject to ramp metering. Recent studies have indicated the feasibility of this mode. In Los Angeles, an unused lane on a metered on-ramp was painted to indicate only car-pool access (Fig. 1)(3). In Minneapolis, following a proposal first developed by the Texas Transportation Institute, special ramps were constructed to allow preferential access to buses (Fig. 2)(6, 10).

This paper develops a ramp-metering design technique appropriate for situations in which preferential access is provided at every ramp by modifying a ramp-metering design technique developed by Payne and Thompson (9). One important aspect of the evaluation provided by this technique is the prediction of ramp waiting times for nonpriority vehicles. Because time saved not waiting in an on-ramp queue may serve as an impetus to form a car pool or to use a bus, identifying ramp waiting times is of interest.

METHODOLOGY

The design of a priority-access ramp-metering plan is approached by optimizing the performance of the traffic pattern generated by a ramp-metering plan. This involves a traffic model in which volumes are taken to be constant over time slices and route selections are made by traffic assignment. The performance measure employed is the total passenger travel time within each time slice.

Freeway Corridor Model

The freeway corridor model that we used was developed by Payne and Thompson (9). It is composed of a series of freeway links and parallel street links connected at interchanges. The network of surface streets between interchanges is aggregated into 1 equivalent street link. This aggregation of flows on the street link is the average of the traffic conditions as seen from the freeway on-ramps. Although some detail is lost, our concern is the relationship between neighboring street volumes and freeway on-ramp volumes. A portion of this network is shown in Figure 3.

The freeway corridor with N interchanges consists of 2 N nodes, 1 for the freeway

Figure 1. Shared-ramp priority access for car pools.

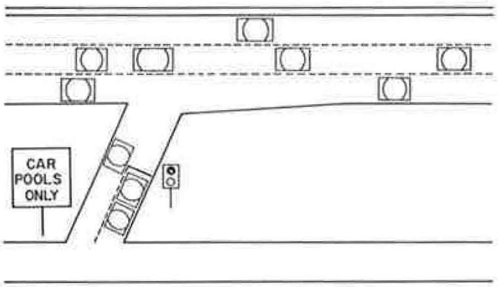
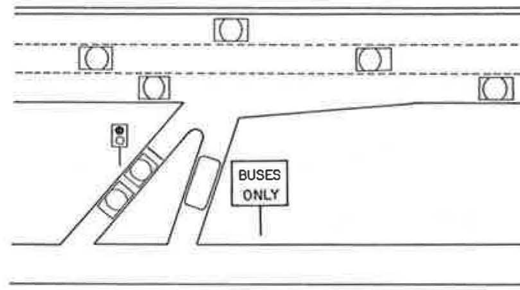


Figure 2. Bus on metered freeway ramp.

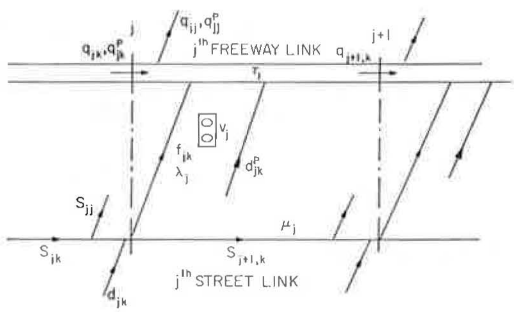


and 1 for the equivalent street. The nodes are connected by N - 1 freeway links and N - 1 street links to a section corresponding to an on-ramp, an off-ramp, or a change in freeway geometry.

Demand volumes are distinguished by origin-destination pairs. Freeway on-ramp and street-link volumes are distinguished by components that represent total flow to a destination; information relating to the origin is not retained.

Priority vehicles are allowed immediate access to the freeway. When they are on the freeway, priority and nonpriority vehicles are assumed to travel at the same speed.

Figure 3. Typical corridor interchange.



We now will put the freeway corridor model into a mathematical format. The following variables are defined for J = 1, 2, ... N and K = 1, 2, ... N, the interchange numbers:

q_{JK} = volume of nonpriority vehicles passing J on the freeway destined for the off-ramp at K ($K \geq J$),
q^P_{JK} = volume of priority vehicles passing J on the freeway destined for the off-ramp at K ($K \geq J$),
f_{JK} = volume of nonpriority vehicles entering the freeway at J destined for the off-ramp at K ($K \geq J$),
d_{JK} = volume of nonpriority vehicles entering the freeway corridor at J destined for K on a surface street ($K \geq J$), and
d^P_{JK} = volume of priority vehicles entering the freeway corridor at J destined for K on a surface street ($K < J$).

We assume that a stationary traffic pattern exists and that flows do not vary in time. Thus there is no storage in a link. By conservation of flow, the flow into a node must be equal to the flow out of a node. For priority vehicles

$$q^P_{J+1,K} = q^P_{JK} + d^P_{JK}, \quad 1 \leq J < K \leq N$$

For nonpriority vehicles

$$q_{J+1,K} = q_{JK} + f_{JK}, \quad 1 \leq J < K \leq N$$

A capacity constraint is placed on the flow in each freeway link, and we define the following variables:

$$q_J = \sum_{K=J}^{N} q_{JK} = \text{total volume of nonpriority vehicles passing J on the freeway, and}$$

$$q_J^P = \sum^{N} q_{JK}^P = \text{total volume of priority vehicles passing J on the freeway.}$$

For each link in the freeway, the capacity constraint is

$$q_J + q_J^P \leq C_J, \quad J = 2, \ldots, N$$

In a similar manner for the street link, we define the following variables:

S_{JK} = volume of vehicles on the street approaching J that are destined for K $(K \geq J)$,

$\tilde{d}_{JK} = d_{JK} + S_{JK}$ = total volume or accumulated demand of nonpriority vehicles on the street at J destined for K $(K \geq J)$,

$S_J = \sum_{K=J}^{N} S_{JK}$ = total volume of nonpriority vehicles on street at J.

For the street network

$$S_{J+1,K} = S_{JK} + d_{JK}, \quad 1 \leq J < K \leq N$$

$$= \tilde{d}_{JK} - f_{JK}$$

<u>Traffic Assignment Algorithm</u>

To determine the set of ramp-metering rates that minimize the performance measure in the freeway corridor, one must determine the resulting traffic pattern for a set of fixed ramp-metering rates. Priority vehicles are assumed to enter at the first available on-ramp so that there will be no difficulty in identifying their routes. Traffic assignment is used to determine the remaining drivers' route choices. The method we used was developed by Payne and Thompson (9) and is similar to that developed by Yagar (13) in that each provides for a queue at each on-ramp. In addition, the method we used produces a traffic pattern that is user optimized. A user-optimized traffic pattern is consistent with Wardrop's first principle (12) that states that, "Journey times on all the routes actually used are equal, and less than those which would be experienced by a single vehicle on any unused route." Under fixed ramp metering, a queue of vehicles will result if the ramp-metering rate is less than or equal to the vehicle arrival rate. If the ramp-metering rate is less than the vehicle arrival rate, the queue will grow extremely large and cause vehicles in the queue to have long delays. When the ramp-metering rate is equal to the vehicle arrival rate, the queue will remain at a fixed length. When the ramp-metering rate is greater than the vehicle arrival rate, there will be no queue. After drivers experiment with alternate routes and when an equilibrium is established, the vehicle arrival rate at the freeway on-ramp will be less than or equal to the ramp-metering rate. Accumulated demand, which is the sum of new demand and existing street traffic in excess of the ramp-metering rate, will be diverted by a street route to the next downstream interchange where drivers may be allowed access to the freeway or diverted again to the street.

In allocating accumulated demand to the freeway on-ramp, one diverts first the vehicles making the shortest trips. Choosing component ramp volumes on this basis is done because drivers having the farthest to travel are more likely to wait in the queue. Allocating accumulated demand to the on-ramp is continued until the ramp-metering rate equals demand. If demand does not exceed the ramp-metering rate, then all accumulated demand is allowed to enter the freeway. If this allocation is performed at interchange J, there exists an interchange number, $\ell(J)$, downstream of interchange J, such that traffic destined for an interchange upstream of $\ell(J)$ will be diverted to the street link, and all traffic destined for an interchange downstream of $\ell(J)$ will be allocated to the freeway on-ramp (Fig. 4).

The traffic assignment algorithm is divided into 3 parts. The first part selects the component ramp volumes, f_{JK}. The choice of component ramp volumes is based on the fact that drivers making the shortest journey will be diverted first. The second part computes link travel times and checks to determine that the freeway capacities (reduced by the volumes of priority vehicles) have not been exceeded. The third part computes ramp queue lengths and travel times.

On-ramp queue lengths are determined by the solution of an equilibrium equation. To facilitate explanation of this equation, we introduce the following variables:

θ_J = time required to cross the jth on-ramp,
λ_J = queue length at the jth on-ramp,
V_J = fixed ramp-metering rate at the jth on-ramp,
τ_{JK} = freeway travel time from interchange J to K (K > J),
t_{JK} = total travel time from interchange J to K on the best route comprised of street and freeway links (K > J), and
μ_J = travel time over street-link J.

The equilibrium equation for interchange J is

$$\theta_J + \lambda_J/V_J + \tau_{J,\ell(J)} = t_{J+1,\ell(J)} + \mu_J$$

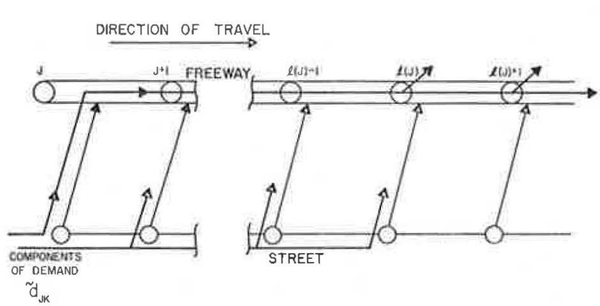
where

$\theta_J + \lambda_J/V_J + \tau_{J,\ell(J)}$ = travel time to $\ell(J)$ when on-ramp J is used, and
$t_{J+1,\ell(J)} + \mu_J$ = travel time to $\ell(J)$ when street-link J is used.

This equation embodies Wardrop's first principle (12). Because drivers traveling to $\ell(J)$ may travel over either of 2 routes, the respective travel times must be the same. This equation is solved for λ_J.

The complete algorithm (9) involves a recursive computation of λ_J, τ_{JK}, and t_{JK} that starts at the downstream end of the corridor and moves to the upstream end.

Figure 4. Bifurcation of ramp demand.



Optimal Allocation

The allocation problem is the determination of the set of ramp-metering rates that minimize total passenger travel-time rate in the freeway corridor. The total passenger travel-time rate is defined in terms of the following variables and those defined earlier.

$\tau_J = \tau_J(q_{J+1} + q^P_{J+1})$ = travel time across freeway link J, a function of link volume,
ρ = average occupancy of nonpriority vehicles, and
ρ^P = average occupancy of priority vehicles.

Total passenger travel-time rate is formulated as

$$\sum_{J=1}^{N-1} (\rho^P q^P_{J+1} \tau_J) + \rho(q_{J+1} \tau_J + S_{J+1} \mu_J + \lambda_J)$$

where

$\rho^P q^P_{J+1} \tau_J$ = passenger travel-time rate in priority vehicles, and
$\rho(q_{J+1} \tau_J + S_{J+1} \mu_J + \lambda_J)$ = passenger travel-time rate in nonpriority vehicles.

Passenger travel-time rate in nonpriority vehicles consists of vehicle travel time on streets, the freeway, and the on-ramp queue. Because priority vehicles are allowed access to the freeway with no wait, no travel time is associated with queues or streets.

If we define the passenger travel-time rate at interchange J as $g_J(q^P_{J+1}, q_{J+1}, S_{J+1}, \lambda_J)$, then the optimal allocation problem would be formulated as

$$\underset{V_1, \ldots, V_{N-1}}{\text{Min}} \sum_{J=1}^{N-1} g_J(q^P_{J+1}, q_{J+1}, S_{J+1}, \lambda_J)$$

subject to the following constraints:

$$q_{J+1} + q^P_{J+1} \leq C_{J+1}, \quad J = 1, \ldots, N-1$$

$$0 \leq V_{JMIN} \leq V_J \leq V_{JMAX}, \quad J = 1, \ldots, N-1$$

The 4 variables in the performance measure are q^P_J, the total volume of priority vehicles in freeway link J; q_J, the total volume of nonpriority vehicles in freeway link J; S_J; and λ_J. q_{J+1}, S_{J+1}, and λ_J are determined by initial freeway and street flow and the set of ramp-metering rates through the traffic assignment algorithm. q^P_{J+1} does not vary with the ramp-metering rate because priority vehicles are allowed direct access to the freeway. q_{J+1}, S_{J+1}, and λ_J can be viewed as state variables dependent on the set of control variables, ramp-metering rates, and upstream corridor conditions. There is a constraint on total freeway volume in each link, and there is an upper and lower bound on each ramp-metering rate. The constraint on freeway capacity is a function of freeway design. There is no constraint on total street-link volume because we assume that capacity is infinite.

For a problem involving 10 or more interchanges, direct optimization (8) would impose a heavy computational burden. One approach to the solution is to formulate the allocation problem as a dynamic programming problem (1). However, this approach is infeasible because of the large number of states. The number of state variables to be considered at a given interchange is equal to the number of downstream interchanges plus 1. A computationally effective compromise developed by Payne and Thompson (9) is to take a suboptimal approach. This greatly reduces the number of states to be considered.

The suboptimal approach considers discrete levels of ramp-metering rates, each of which bifurcates accumulated demand at an interchange between the street and the freeway link at a level corresponding to 1 of the downstream interchanges. For instance, at interchange J, there are N - J possible metering rates as follows:

$$\tilde{d}_{JN},\ \tilde{d}_{JN} + \tilde{d}_{J,N-1}, \ldots, \sum_{k=J+1}^{N} \tilde{d}_{JK}$$

Only those rates that meet minimum and maximum metering-rate constraints are considered.

Details of the algorithm that provides optimal bifurcation metering rates are given elsewhere (9).

DESIGN EXAMPLES

Freeway Location and Geometry

The segment of freeway used in this example is part of northbound I-405 (San Diego Freeway) in Los Angeles County, California, from the Vermont Avenue on-ramp to a point that is just upstream of the eastbound Imperial Highway off-ramp. Figure 5 shows this section of the freeway.

Fixed ramp metering has been used on this freeway segment since 1972 (5). For our purposes, the freeway segment is divided into 10 links. The upstream boundary of each link corresponds to the first off-ramp in the interchange. Link 6, however, has no off-ramp. Links 8, 9, and 10 have 2 on-ramps, and link 10 has 2 off-ramps. The on-ramps and off-ramps in these links correspond to eastbound and westbound street-traffic on-ramps and off-ramps near interchanges 9 and 10 and northbound and southbound street-traffic on-ramps and off-ramps near interchange 8.

Table 1 gives some additional information about the freeway geometry. Each freeway link has 4 lanes except for link 9, which has 5. Freeway capacity is assumed to be 2,025 vehicles per hour (vph) per lane. Link 9's capacity, which is reduced, is 1,800 vph per lane. This lower value was chosen because we assumed that the higher freeway capacity in the other links could not be maintained where merging was anticipated. The upper bound on on-ramp rates is 900 vph per lane.

Traffic Engineering Data

The volume and density data used to estimate the speed-volume relationship and origin-destination pairs were derived from 1-min averages of loop-sensor data. The loop-sensor data for this segment of freeway for April 23, 1974, from 6:30 to 9:00 a.m. were obtained from the California Department of Transportation.

A parabolic relationship between speed and volume was used. The region of the curve representing uncongested flow was used for subsequent computations of speed from volume. A freeway loop-sensor station nearest the upstream boundary of a link was used to derive the speed-volume relationship for that link by a least squares fit. A sample of the least squares fit of the speed-volume relationship is shown in Figure 6.

The volume data for the upstream boundary of the freeway segment and the volume

Figure 5. Freeway segment for design example.

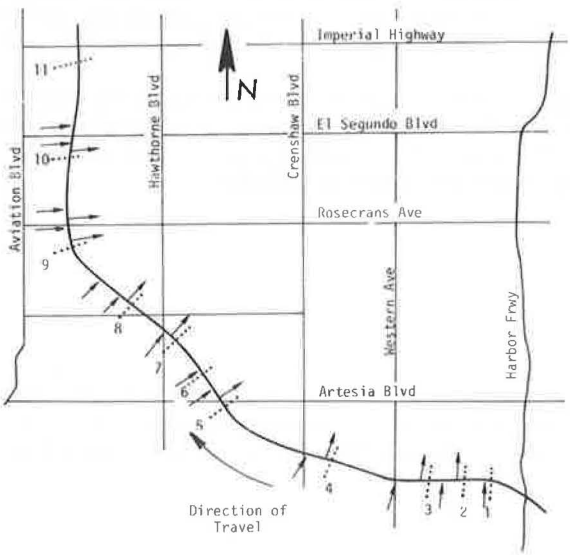
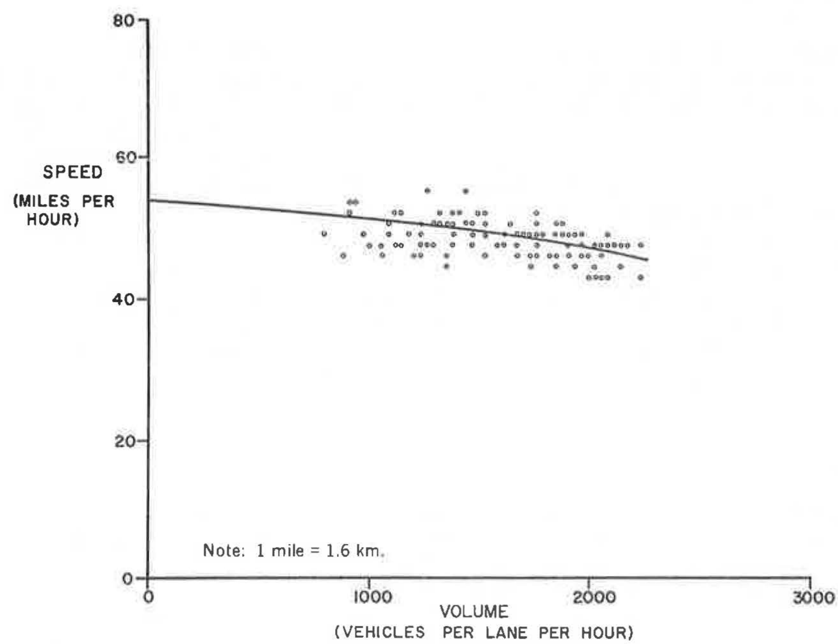


Table 1. Corridor geometry.

Interchange Link	Link Length (miles)
1. Vermont Avenue	0.408
2. Normandie Avenue	0.542
3. Western Avenue	1.051
4. Crenshaw Boulevard	1.294
5. Artesia Boulevard	0.431
6. Redondo Beach Boulevard	0.428
7. Hawthorne Boulevard	0.647
8. Inglewood Boulevard	0.983
9. Rosecrans Avenue	0.896
10. El Segundo Boulevard	0.790

Note: 1 mile = 1.6 km.

Figure 6. Speed-volume relationship.



data for the on-ramp and off-ramp volumes were averaged over 15-min intervals starting at 6:30 a.m. Origin-destination data, which were compatible with observed data, were estimated.

Comparison to Present Design

The traffic assignment algorithm was used to determine the traffic pattern and performance measure in the corridor that corresponded to current ramp-metering design and estimated origin-destination data. A new ramp-metering design was determined by the optimal allocation algorithm for the same origin-destination data. In both applications of the methodologies explained previously, identical speed-volume relationships and freeway geometry data for each link were used as inputs to the 2 algorithms. With the 2 algorithms, a comparison was made of the difference between the current design and optimal design from 6:30 to 8:00 a.m. Comparisons of current ramp-metering design and optimal ramp-metering design are given in Tables 2 and 3. The comparison shows that there is a difference in the 2 designs. The optimal allocation algorithm tends to fill an unused portion of the freeway. Note, however, that the current design was developed from a different set of origin-destination data. Table 4 gives a comparison of the performance measure for the 2 different designs. The difference in the total passenger travel time of the 2 designs is not significant, which indicates that the optimal ramp-metering design produces only marginal changes in freeway corridor performance. An Appendix[1] includes additional comparisons of link speeds, travel times, and ramp queue lengths.

Table 2. On-ramp metering rates for links 1 through 5.

	Link 1		Link 2		Link 3		Link 4		Link 5	
Time	Current Design	Optimal Design	Current Design	Optimal Design	Current Design	Optimal Design	Current Design	Optimal Design	Current Design	Optimal Design
6:30 a.m.	180	172	180	172	180	195	240	424	180	551
6:45 a.m.	180	475	180	188	360	212	240	472	240	548
7:00 a.m.	140	470	220	176	300	244	360	632	240	180
7:15 a.m.	200	551	140	172	260	280	580	508	480	208
7:30 a.m.	120	596	100	172	260	180	520	508	460	256
7:45 a.m.	140	184	120	338	220	176	540	384	140	500

Note: Values are in vehicles per hour.

Table 3. On-ramp metering rates for links 6 through 10.

	Link 6		Link 7		Link 8		Link 9		Link 10	
Time	Current Design	Optimal Design	Current Design	Optimal Design	Current Design	Optimal Design	Current Design	Optimal Design	Current Design	Optimal Design
6:30 a.m.	180	216	180	552	720	551	1,580	1,273	1,000	320
6:45 a.m.	180	236	180	581	720	596	720	1,388	960	344
7:00 a.m.	180	196	180	368	540	632	1,440	1,276	1,000	352
7:15 a.m.	560	208	350[a]	344	540	1,177	1,280	1,368	980	368
7:30 a.m.	560	188	350[a]	342	840	1,454	1,360	1,600	1,240	400
7:45 a.m.	500	352	660	360	1,040	568	840	1,760	800	360

Note: Values are in vehicles per hour.

[a]Design on ramp meter changed to reflect observed values.

[1] The original manuscript of this paper included an appendix. This appendix is available in Xerox form at cost of reproduction and handling from the Transportation Research Board. When ordering, refer to XS-55, Transportation Research Record 533.

Table 4. Total passenger travel-time rates.

Time	Current Design (passenger hours)	Optimal Design (passenger hours)
6:30 to 6:45 a.m.	2,055	2,047
6:45 to 7:00 a.m.	2,100	2,059
7:00 to 7:15 a.m.	2,201	2,183
7:15 to 7:30 a.m.	2,061	2,035
7:30 to 7:45 a.m.	2,135	2,100
7:45 to 8:00 a.m.	2,245	2,233
6:30 to 8:00 a.m.	3,199	3,164

Ramp-Metering Plans Under Modal Shift

An analysis of the effect of modal shift on total passenger travel time within the corridor was performed by using the optimal allocation algorithm. Modal shift is the difference in the percentage of passengers using priority vehicles from the initial distribution of passengers in vehicles. The analysis is performed to study the differential between the passenger travel times for priority and nonpriority vehicles, which is of particular interest because this difference may act as an incentive to drivers to shift to the priority-vehicle mode. It is assumed that the results are measured after a modal shift has occurred.

The distribution of vehicle occupancy is as follows:

Passengers per Vehicle	Percentage of Vehicles
1	80
2	15
3	5

The average occupancy of this distribution is 1.25 passengers per vehicle. It can be seen from this distribution that 12 percent of the passengers in the freeway corridor would be given preferential treatment, if no modal shift occurred and preferential access were allowed to vehicles with 3 or more passengers. The method for calculating the reduction of vehicles by a change in the mode of travel of passengers for a nonpriority vehicle to a priority vehicle is as follows. Five new priority vehicles containing 3 passengers are created from 3 nonpriority vehicles containing 2 passengers and 9 nonpriority vehicles containing 1 passenger. The change in mode of travel is done uniformly throughout the corridor. The reduction in vehicular demand occurs both for upstream components of total freeway volume and for components of total on-ramp demand. During the analysis of different levels of modal shift, the passenger demand from an origin to a destination remains constant. However, as the modal shift increases, demand, in terms of vehicles, decreases.

The optimal allocation algorithm was used to determine the ramp-metering plan and the resultant traffic pattern for different levels of modal shift. Figure 7 shows a comparison of passenger travel times for increasing modal shifts. The greatest decrease in passenger travel time occurs when 24 to 36 percent of the passengers are in priority vehicles.

If we reallocate passengers by the method previously described, a 12 percent increase in passengers using priority vehicles would produce a 7 percent decrease in total corridor demand. Estimated demand in the traffic corridor would exceed freeway capacity by 7 to 14 percent because on-ramp queues are reduced significantly after a modal shift of 12 percent. Table 5 gives queue waiting time for different levels of modal shift. It is apparent from Table 5 that there is little incentive after a modal shift of 12 percent. However, this level of modal shift is sufficient to reduce demand in the corridor to freeway capacity.

A sensitivity analysis was performed to determine the effect of overestimation or underestimation of modal shift on total passenger travel time in the corridor. A set of optimal ramp-metering designs for a specific modal shift were chosen, and the traffic pattern and performance measure for different modal shifts were determined. Table 6 gives the results of the sensitivity analysis.

It can be seen from the data given in Table 6 that, if modal shift is overestimated, if optimal ramp-metering design is based on a modal shift that is greater than actual

Figure 7. Passenger travel time, 6:30 to 8:00 a.m.

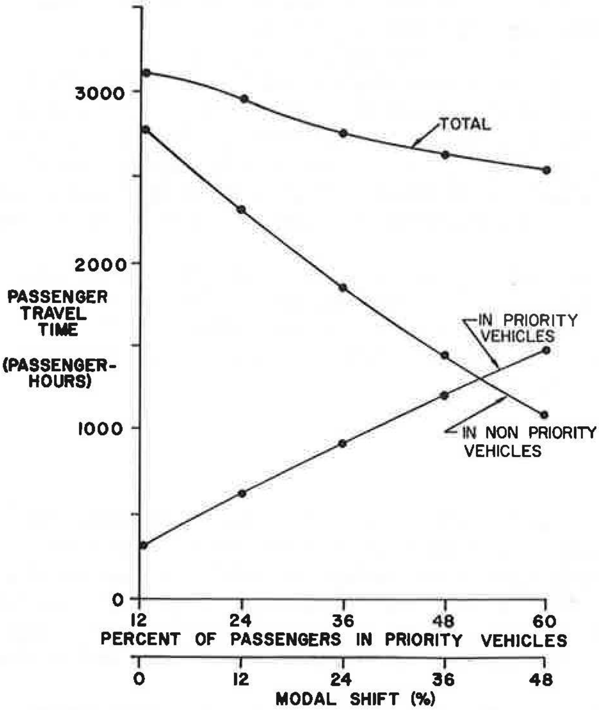


Table 5. On-ramp waiting times.

	Wait (min)									
Level	Link 1	Link 2	Link 3	Link 4	Link 5	Link 6	Link 7	Link 8	Link 9	Link 10
No priority access	3.3	2.8	2.6	2.2	1.3	1.3	1.1	0.6	0	0
Priority access										
No modal shift	3.2	2.8	2.6	2.1	1.3	1.3	1.1	0.6	0.1	0
12 percent modal shift	0.4	2.8	2.6	2.2	1.3	1.0	0.7	0.3	0.1	0
24 percent modal shift	0	2.4	2.2	1.6	0.8	0.4	0.2	0	0.1	0
36 percent modal shift	0	0	0	0	0	0	0	0	0	0

Table 6. Sensitivity of performance to estimates of modal shift.

	Ramp-Plan Modal Shift			
Modal Shift	None	12 Percent	24 Percent	36 Percent
None	2,012	—[a]	—[a]	—[a]
12 percent	1,977	1,857	—[a]	1,994
24 percent	1,844	1,851	1,777	1,929
36 percent	1,870	1,800	1,798	1,713

Note: Values are in passenger hours.

[a]Traffic demand exceeds freeway capacity.

modal shift, freeway links may be congested. The greater the difference is between actual modal shift and modal shift assumed in developing the design, the greater the degradation is in performance when modal shift is underestimated.

When modal shift is overestimated, the freeway segment may become congested. The optimal ramp-metering design used is based on a demand that is less than actual demand. In this case, ramp-metering rates are chosen that allow more vehicles on the freeway because upstream freeway volume is reduced. When modal shift is underestimated, the freeway segment is underused. The optimal ramp-metering design allows fewer vehicles on the freeway because upstream freeway volume is assumed to be larger than actual volume. As a consequence, vehicles are unnecessarily denied access to the freeway from on-ramps.

The sensitivity analysis and comparison of corridor performance when preferential access is presented are based on estimated origin-destination data. The estimated origin-destination data do not encompass latent demand within the corridor. If preferential access is allowed, reduced corridor demand might attract more vehicles that currently are not entering the freeway corridor.

CONCLUSIONS

In this paper, techniques for evaluating and designing priority-access ramp-metering plans have been presented. The design technique presented generates a ramp-metering plan that minimizes total passenger travel time in a freeway corridor and predicts the traffic pattern, including ramp queues, that would result. These techniques are limited to use in situations in which every on-ramp has priority access. To extend these techniques to situations in which only certain ramps have priority access would require new methodological development.

The presence of ramp queues might serve to induce drivers without passengers to form or join car pools or to use buses. The methodology presented here includes a prediction of on-ramp waiting times so that one might assess this factor as a motivation for a modal shift. In the example presented here, there was little incentive to form additional car pools after a modal shift of about 12 percent.

REFERENCES

1. R. E. Bellman. Dynamic Programming. Princeton Univ. Press, New Jersey, 1957.
2. Freeway Lanes for High-Occupancy Vehicles. California Department of Transportation, 3d Annual Progress Rept., Dec. 1973.
3. R. G. B. Goodell. Preferential Access for Multi-Occupant Vehicles at Metered On-Ramps. Proc., Greater Los Angeles Area Symposium, 1972-1973, Los Angeles Council of Engineers and Scientists, American Institute of Aeronautics and Astronautics, Vol. 12.
4. D. R. Gordon. Preliminary Ramp Control Plan, Northbound San Diego Freeway From Route 11 to El Segundo Boulevard. California Division of Highways, Rept. 70-9, Nov. 1970.
5. D. R. Gordon. A Preliminary Evaluation of Ramp Control on the San Diego Freeway in Los Angeles. California Division of Highways, Rept. 72-20, Dec. 1972.
6. Evaluation Operation Manual for the I-35W Urban Council Demonstration Project. Metropolitan Council, St. Paul, Minn., Oct. 1973.
7. R. D. Minister, L. P. Lew, K. Ovaici, and A. D. May. A Computer Simulation Model for Evaluating Priority Operations on Freeways. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, June 1972, p. 317.
8. W. C. Mylander, R. L. Holmes, and G. P. McCormick. A Guide to SUMT-Version 4: The Computer Program Implementing the Sequential Unconstrained Minimization Technique for Nonlinear Programming. Research Analysis Corp., RAC-P-63, Oct. 1971.
9. H. J. Payne and W. A. Thompson. Allocation of Freeway Ramp Metering Volumes

to Optimize Corridor Performance. Institute of Electrical and Electronics Engineers, Transactions on Automatic Control, Vol. AC-19, No. 3, June 1974, pp. 177-186.

10. A System to Facilitate Bus Rapid Transit on Urban Freeways. Texas Transportation Institute, Texas A&M Univ., Dec. 1968.
11. Preferential Treatment of High Occupancy Vehicles. Federal Highway Administration, U.S. Department of Transportation, April 1974.
12. J. G. Wardrop. Some Theoretical Aspects of Road Traffic Research. Proc., Institute of Civil Engineers, 1952, pp. 325-363.
13. S. Yagar. Dynamic Traffic Assignment by Individual Path Minimization and Queuing. Transportation Research, Vol. 5, Aug. 1971, pp. 179-196.

DEVELOPING FREEWAY PRIORITY-ENTRY-CONTROL STRATEGIES

Khosrow Ovaici, Roger F. Teal, James K. Ray, and Adolf D. May,
Institute of Transportation and Traffic Engineering,
University of California, Berkeley

This paper describes the development and application of an analytical procedure for priority-entry-control strategies at freeway ramps. Vehicles with different numbers of occupants arriving at an on-ramp are differentiated, and those vehicles with more occupants are given priority entry onto the freeway. The 2 primary objectives of the priority-control strategy are to maximize either the number of persons served or the number of passenger-miles traveled. The primary constraint is that the vehicular demand for each freeway section not exceed the vehicular capacity of that freeway section. Additional constraints such as maximum and minimum metering rates can be specified. The analytical procedure encompasses 2 models. The first is a simulation model that is deterministic and macroscopic and predicts freeway-traffic performance as a function of freeway design and allowable ramp inflows. The second is a decision model that has a linear programming formulation and selects a control strategy that meets specified objectives and constraints. The simulation model has been validated under field conditions, and the predicted traffic performance compares favorably to actual, measured traffic performance. The 2 models have been integrated and computerized, and the composite model has been applied to the East Bayshore Freeway in the San Francisco Bay area, the Santa Monica Freeway in Los Angeles, and the Long Island Expressway to demonstrate its application and to provide the California Department of Transportation and the New York State Department of Transportation with results that could be considered for possible implementation. A series of investigations were undertaken with the computerized model to determine the sensitivity of the overall measures of effectiveness to practical constraints and to consider the consequences of such control strategies on changing the traffic-demand pattern and passenger-car occupancy distributions.

•MANY urban freeway systems have congested segments during peak traffic periods. When this congestion occurs, other portions of the freeway system can be adversely affected—productivity is reduced [fewer passenger-miles (kilometers) of travel]; level of service is reduced (greater passenger hours of travel); and accidents, pollution, and energy consumption are increased. Congestion occurs when vehicular demands exceed roadway capacities. And congestion often occurs when adjacent time periods and parallel alternate routes are not congested. Two possible solutions are available to eliminate congestion: increase roadway capacity or reduce vehicular demand.

During the past 3 decades in the United States, congestion has been reduced primarily by increasing the roadway capacities of the freeway system. When the freeway system was not extensive, and constraints such as the limitations on the use of urban land and the requirements of environmental protection were not so restrictive, this was an effective approach. Although increasing roadway capacities has resulted in higher levels of service, it also has had the unfortunate consequence of encouraging, and, in many cases creating, even greater vehicular demands. And there has been increased concern about the extensive urban land required for vehicular movement and parking, the increasing number of accidents and, recently, air pollution and energy consumption.

During the last decade attention has begun to shift to the other solution—the reduction of vehicular demand on congested roadways during peak periods. Control on entry to the freeway and priority lanes for multipassenger vehicles has been employed to reduce congestion. The former reduces vehicular demand by spreading excess vehicular demand to other time periods or other routes or both. The latter modifies vehicular demand by encouraging car pools and bus travel. The main thrust of this paper is to model and evaluate an implementation strategy that integrates entry control and priority treatment into a priority-entry-control system.

Consider a directional freeway that has a number of entry points. At each entry point, vehicular demand is separated into 2 traffic streams—priority vehicles and nonpriority vehicles. A priority vehicle contains n or more passengers; such vehicles would be permitted to enter the freeway without stopping. Nonpriority vehicles would pass through a queuing process and would be permitted to enter the freeway on a space-available basis. Undoubtedly, the implementation of such a priority-entry-control system will require careful consideration of driver education, enforcement, and traffic engineering. However, an experiment has been under way in Los Angeles for the past year at 2 ramps, and the results indicate that such a priority-entry-control system is operationally feasible and can be satisfactorily implemented in the field (1, 2).

LINEAR PROGRAMMING FORMULATION FOR PRIORITY-ENTRY CONTROL

Many decision problems are formulated now as mathematical programming problems, requiring the maximization or minimization of an objective function subject to constraints. Application of linear programming techniques to the problem of freeway on-ramp control was first demonstrated by Wattleworth in 1964 (3). Later work was done by Goolsby, Merrell, and McCasland (4); Brewer et al. (5); and Wang and May (6). A priority-entry-control algorithm using the linear programming upper-bounding method was first formulated by Ovaici and May (7). This paper is an extension and application of this work.

Basic Priority-Entry Formulation

The study section of the freeway is divided into homogeneous subsections that exhibit properties of constant capacity and demand over their length.

A basic priority-entry-control strategy has been developed in the form of a linear programming problem that has an objective function of maximizing the number of persons served and a primary constraint that the demand for each freeway section not exceed the capacity of that freeway section.

Maximize

$$\sum_{i=1}^{n} (X_{i1} + 2X_{i2} + 3X_{i3} + 4X_{i4} + 5X_{i5} + b_i \cdot X_{i6}) \tag{1}$$

subject to

$$\sum_{i=1}^{n} (F_{i1\ell} \cdot X_{i1} + F_{i2\ell} \cdot X_{i2} + , \ldots , + F_{i5\ell} \cdot X_{i5} + F_{i6\ell} \cdot e \cdot X_{i6}) \leq C_\ell \tag{2}$$

for $\ell = 1, 2, \ldots, P$;

$$X_{ik} \leq D_{ik} \tag{3}$$

for $i = 1, 2, \ldots, n$; and $k = 1, 2, \ldots, 6$;

$$X_{ik} \geq 0 \tag{4}$$

for i = 1, 2, ..., n; and k = 1, 2, ..., 6, where

X_{ik} = input flow rate at on-ramp i, for traffic with passenger occupancy k(k = 1, 2, ..., 5),
X_{i6} = input flow rate at on-ramp i, for buses,
n = number of on-ramps,
p = number of freeway subsections,
$D_{ik} = \sum_{j=1}^{m} d_{ijk}$ = traffic demand rate for on-ramp i with passenger occupancy k(k = 1, 2, ..., 5),
m = number of off-ramps,
d_{ijk} = traffic demand rate from on-ramp i to off-ramp j with passenger occupancy k(k = 1, 2, ..., 5),
$D_{i6} = \sum_{j=1}^{m} d_{ij6}$ = bus demand rate for on-ramp i,
d_{ij6} = bus demand rate from on-ramp i to off-ramp j,
$F_{ik\ell}$ = fraction of traffic X_{ik} that passes through subsection ℓ,
C_ℓ = capacity of subsection ℓ,
b_i = bus occupancy at on-ramp i, and
e = bus equivalency factor.

In this formulation, on-ramps, off-ramps, and freeway subsections are numbered from upstream to downstream. Equation 1 states that the objective of the control is to maximize the total passenger input rate from all on-ramps. Equation 2 is the capacity constraint that total vehicular demand for any subsection should not exceed its capacity. Equations 3 and 4 are demand and nonnegativity constraints respectively.

To calculate coefficient $F_{ik\ell}$, the origin-destination (O-D) patterns of all classes (class k for k = 1, 2, ..., 6) of vehicles (d_{ijk}) must be available. If d_{ijk} is not available, $F_{ik\ell}$ can be estimated from $O\text{-}D_{ij}$ (where $O\text{-}D_{ij}$ = origin-destination pattern of all classes of vehicles), assuming all classes of vehicles have a similar O-D pattern. In this case, the percentage of each class of vehicles (based on passenger occupancy) at each on-ramp must be given. Then

$$D_{ik} = POC_{ik} \sum_{j=1}^{m} OD_{ij}$$

where POC_{ik} = percentage of class k vehicles for on-ramp i.

Because the objective function (Eq. 1) and constraints (Eqs. 2 and 3) are linear, this problem can be solved by the regular simplex method. But, because of the special structure of the problem, upper-bounding linear programming can be employed, which results in a significant gain in computation efficiency and a reduction in computer memory requirements. By using the upper-bounding method, the size of this linear programming problem (because of its special structure) will be decreased by a ratio of up to 9.

Underlying Assumptions

A number of assumptions are made in order that the linear programming formulations can be applied to real-life problems. These assumptions are that

1. Time can be divided into discrete, equally spaced intervals called time slices;
2. Space (the length of the freeway) can be divided into homogeneous subsections,

each of which exhibits the properties of constant capacity and demand over their lengths;

3. Within a given time slice, traffic demands remain constant and do not fluctuate over that time slice;
4. When traffic demands are loaded onto the freeway, demands propagate downstream instantaneously unless there are capacity constraints; and
5. Traffic diverted from one on-ramp will not enter other on-ramps.

Extension of Basic Priority-Entry Formulation

To be able to solve a wide variety of real-life problems the basic, priority-entry formulation has been extended to encompass

1. Additional objective functions,
2. Metering rate limits,
3. Operational control constraints,
4. Main-line input fluctuation,
5. Capacity buffer and level-of-service constraint,
6. Short-trip formulation, and
7. Multi-time-slice control.

Additional Objective Functions

In the basic priority-entry formulation, maximizing the number of persons served was chosen as the objective function. In this section, 3 other objective functions will be developed.

The first is maximizing total passenger-miles (kilometers) of travel as follows. Maximize

$$\sum_{\ell=1}^{n} (\ell_{11}X_{11} + 2\ell_{12}X_{12} + \ldots + 5\ell_{15}X_{15} + b \cdot \ell_{16} \cdot X_{16}) \tag{5}$$

where

ℓ_{ik} = average trip length of traffic with passenger occupancy $k(k = 1, 2, \ldots, 5)$ for on-ramp i, and
ℓ_{i6} = average trip length of buses for on-ramp i.

The second is maximizing total number of vehicles served. Maximize

$$\sum_{i=1}^{n} \sum_{k=1}^{6} X_{ik} \tag{6}$$

The third is maximizing total vehicle miles of travel. Maximize

$$\sum_{i=1}^{n} \sum_{k=1}^{6} \ell_{ik}X_{ik} \tag{7}$$

Thus the model includes 4 optional objective functions, namely maximizing vehicle input, maximizing vehicle miles (kilometers) of travel, maximizing passenger input, and maximizing passenger-miles (kilometers) of travel. The first 2 objective functions are for control on a vehicle basis (all vehicles are treated the same regardless of passenger occupancy), and the last 2 objectives are for control on a passenger basis (vehicles with different occupancies are differentiated, and those vehicles with higher occupancies will be given priority entry onto the freeway).

Metering Rate Limits

Maximum and minimum metering rate limits can be entered as constraints and expressed mathematically as follows:

$$\sum_{k=1}^{6} X_{ik} \leq M_i \tag{8}$$

for i = 1, 2, . . . , n;

$$\sum_{k=1}^{6} X_{ik} \geq m_i \tag{9}$$

for i = 1, 2, . . . , n, where M_i and m_i are respectively the maximum and minimum metering rates for on-ramp i. The minimum metering rate is necessary to prevent excessive driver violation at the on-ramp or to prevent the ramp queue from backing up onto the arterial streets or both. The maximum metering rate for nonpriority vehicles may be required because of the geometric design of the on-ramp and the hardware capacity of the metering system.

Operational Control Constraints

Operational control constraints may be added for any combination of on-ramps. The various options are no control (at on-ramp i), that is,

$$X_{ik} = D_{ik} \tag{10}$$

for k = 1, 2, . . . , 6; automobiles only, that is,

$$X_{i6} = 0 \tag{11}$$

priority vehicles only, that is,

$$X_{i1} = X_{i2} = \ldots = X_{i,f-1} = 0 \tag{12}$$

buses only, that is,

$$X_{i1} = X_{i2} = X_{i3} = X_{i4} = X_{i5} = 0 \tag{13}$$

and ramp closed, that is,

$$X_{i1} = X_{i2} = \ldots = X_{i6} = 0 \tag{14}$$

where f = priority cutoff level.

Sometimes for practical reasons it may be necessary to have a preset priority cutoff level for some or all on-ramps. This can be implemented by adding the following constraints:

$$X_{ik} = D_{ik} \tag{15}$$

for k = h, h + 1, . . ., 6 for some or all i where h = preset, priority cutoff level.

Main-Line Input Fluctuation

One critical weakness in fixed-time control is that the input rate from the main-line upstream point is a variable that cannot be controlled. If mean arrival rates are used in the model to determine optimum priority-control strategy, there is a 50 percent probability that congestion will occur despite the control if one assumes that ramp input rates are uniformly distributed, that the arrival rate of the main-line input is normally distributed, and that the trip pattern is constant. There is also a 50 percent probability that the freeway will be overcontrolled. Congestion, however, is highly undesirable and should be prevented in almost all cases even at the expense of overcontrol.

Let

EV = the expected flow rate of the main-line input,
SD = the standard deviation of the flow, and
DV = the design flow rate to be used in the model.

Then, for a specified confidence $1 - \alpha$, where α is the probability that the observed flow rate is higher than the design flow rate, the design flow rate can be found by

$$DV = EV + U_{1-\alpha}SD \tag{16}$$

where

$$\Phi(U_{1-\alpha}) = 1 - \alpha \tag{17}$$

$U_{1-\alpha}$ is $1 - \alpha$, normal at the 100th percentile, that is,

$$\Phi(U_{1-\alpha}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{U_{1-\alpha}} e^{-t^2/2} dt \tag{18}$$

If one uses DV in the model, the resulting control strategy should be free of recursive congestion (congestion caused by normal roadway and traffic congestion) with a probability of $1 - \alpha$.

Capacity Buffer and Level-of-Service Constraint

In Eq. 2 C_ℓ can be replaced by SV, service volume of subsection ℓ for a given level of service. Travel speed has been selected as the major factor to use in identifying the level of service (8). A second factor—either the ratio of demand volume to capacity or the ratio of service volume to capacity, depending on the particular problem situation—is also used in making this identification (8). In practice, the second factor is referred to as the v/c ratio. A minimum operating speed can be specified to reflect the desired minimum level of service. For a given operating speed versus v/c curve, specifying a maximum v/c value is equivalent to specifying a minimum operating speed.

The capacity buffer can be expressed by either the excess capacity (expected capacity minus allowable volume) or the excess v/c value (1 minus allowable v/c value). Thus Eq. 2 can be replaced by 1 of the following equations:

$$\sum_{i=1}^{n} (F_{1 1 \ell} \cdot X_{11} + F_{12\ell} \cdot X_{12} + \ldots + F_{15\ell} \cdot X_{15} + F_{16\ell} \cdot e \cdot X_{16}) \le C_\ell \; (1 - EVOC) \tag{19}$$

or

$$\sum_{i=1}^{n} (F_{1 1 \ell} \cdot X_{11} + F_{12\ell} \cdot X_{12} + \ldots + F_{15\ell} \cdot X_{15} + F_{16\ell} \cdot e \cdot X_{16}) \le C_\ell - ECA \tag{20}$$

where

EVOC = excess v/c value, and
ECA = excess capacity.

Short-Trip-Diversion Formulation

An alternative linear programming formulation, based on the concept that people taking short trips are more likely to divert to alternative routes than are people taking long trips, is presented in this section. In the basic formulation of the priority-control strategy it is assumed that for each on-ramp the destination pattern before and after priority control is the same; this pattern is reflected by the parameter $F_{ik\ell}$ in Eq. 2, which is computed from the destination pattern before control. This assumption is reasonable if little or no diversion occurs. Little or no diversion occurs when the metering rates are only slightly less than the demand or when there is no suitable alternative route.

In general, when metering rate is less than demand for an on-ramp, a queue will form on the ramp and cause a certain amount of delay to nonpriority vehicles. Some of the vehicles may prefer to use alternative routes. Traffic with better alternative routes or a smaller travel-time difference between the alternative route and the freeway route is likely to divert first. The exact pattern of diversion is undoubtedly stochastic in nature and depends on the actual origin and destination of each trip and on driver characteristics. As an approximation, it is assumed that single-occupancy vehicles with shorter freeway trip lengths will divert proportionally more than will single-occupancy vehicles with longer freeway trip lengths. This assumption can be expressed as follows:

$$Y_{ij} \leq Y_{i,j+1} \tag{21}$$

for all (i, j) where Y_{ij} = percentage of original demand of single-occupancy vehicles from on-ramp i to off-ramp j that is not diverted.

Diverted vehicles are taken from the lowest occupancy class possible; that is, the vehicles to be diverted are removed first from single-occupancy vehicle demand, then from the double-occupancy vehicle demand and so forth until the total number of vehicles to be diverted is satisfied. The single-occupancy vehicles are diverted, if required; Eq. 21 illustrates the pattern of such diversion. The diversion pattern for vehicles with 2 or more occupants will be identical to the vehicles' original demand pattern.

For this alternative formulation, X_{i1} in Eq. 1 will be replaced by

$$X_{i1} \leftarrow \sum_{j=1}^{m} \delta_{ij} \cdot Y_{ij} \cdot d_{ij1}$$

where

δ_{ij} = 1 if j is downstream of i or 0 if j is upstream of i,
m = number of off-ramps, and
d_{ij1} = demand of single-occupancy vehicles from on-ramp i to off-ramp j.

Therefore, the objective function and the capacity constraint shown earlier in Eqs. 1 and 2 become maximized. Maximize

$$\sum_{i=1}^{n} \left[\sum_{j=1}^{m} \delta_{ij} \cdot Y_{ij} \cdot d_{ij1} + 2X_{i2} + 3X_{i3} + 4X_{i4} + 5X_{i5} + b_i X_{i6} \right] \tag{22}$$

subject to

$$\sum_{i=1}^{n}\left[\sum_{j=1}^{m} \gamma_{ij} \cdot Y_{ij} \cdot d_{ij\ell} + F_{i2\ell} \cdot X_{i2} + F_{i3\ell} \cdot X_{i3} + F_{i4\ell} \cdot X_{i4} + F_{i5\ell} \cdot X_{i5} + F_{i6\ell} \cdot e \cdot X_{i6}\right] \le C_\ell \tag{23}$$

for $\ell = 1, 2, \ldots, p$ where $\gamma_{ij} = 1$ if i is upstream of subsection ℓ and j is downstream of subsection ℓ, and 0 otherwise.

The other 2 constraint equations (Eqs. 3 and 4) are replaced by the following 4 constraint equations:

$$Y_{ij} \ge 0 \tag{24}$$

for all (i, j),

$$Y_{ij} \le 1 \tag{25}$$

for all (i, j),

$$X_{ik} \le D_{ik} \tag{26}$$

for k = 2, 3, ..., 6 and all i, and

$$X_{ik} \ge 0 \tag{27}$$

for k = 2, 3, ..., 6 and all i where X_{ik} and Y_{ij} are decision variables.

Multi-Time-Slice Control

The peak period is divided into discrete, equally spaced intervals called time slices. In general, when the metering rate for a time slice is less than the demand for an on-ramp, a queue will form on the ramp and cause a certain amount of delay to nonpriority vehicles. Some of the vehicles will find it more suitable to use alternative routes. People with better alternative routes or a smaller travel-time difference between the alternative route and the freeway are likely to divert first.

Ramp vehicles waiting in the queue at the end of a time slice become, in effect, part of the demand of the following time slice. The length and the trip pattern of the ramp queue under control are functions of the original demand pattern, priority cutoff level, metering rate for nonpriority vehicles, driver behavior, and network configuration.

Priority-control strategies for 2 cases of traffic diversion that will be developed are total diversion control and no diversion control.

Total diversion control requires that all nonpriority trips in excess of the nonpriority metering rate be diverted to arterial streets. This may be a good approximation of freeway corridors with good alternative routes.

No diversion control assumes that no vehicle will divert to arterial streets. In this case the ramp queue is equal to demand minus metering rate. At the end of each time slice the queue is added to the original demand of the next time slice to become the total demand for that time slice. Then the control strategy is developed by using total demand as input to the model. The assumption of no diversion applies to freeway corridors that do not have suitable alternative routes or to situations where the metering rate is only slightly less than demand.

The control strategies developed for total diversion and no diversion cases are extreme. Actual diversion lies between these 2 cases. Further research is now under way in regard to partial diversion.

INTEGRATING LINEAR PROGRAMMING FORMULATION AND FREEWAY MODEL

Purpose of Integration

It is desirable to integrate the linear programming model (decision model) with a freeway model (simulation model) for the following 3 reasons:

1. Some interactions among weaving, merging, diverging capacity, and the selected ramp-control strategy cannot be handled independently;
2. Traffic performance at entrance ramps and along the freeway is dependent on traffic diversion; and
3. Feasibility analysis and refinement of the control strategy require traffic performance information.

Model Structure

The proposed analytical procedure includes 2 models. The first is a deterministic and macroscopic simulation model that predicts freeway performance as a function of freeway design and traffic demand. The second is a decision model that has a linear programming formulation; it selects the control strategy that maximizes the objective function subject to the stated set of constraints.

The freeway model (FREQ3) was developed during a freeway operations study at the Institute of Transportation and Traffic Engineering, University of California, Berkeley (11). This model has been validated under field conditions, and the predicted traffic performance compares very favorably to actual, measured traffic performance.

MODEL COMPUTERIZATION

This section will describe 3 computerized models: PREFO, FREQ3, and FREQ3CP (9).

Decision Model, PREFO

Based on the linear programming formulation previously described, a computer program, PREFO (priority entry at freeway on-ramps), was prepared. The PREFO computer program consists of a main program and 12 subroutines. A wide variety of options are available in the PREFO program that provide the user with a versatile model. Table 1 gives the major available options in the model.

Freeway Model, FREQ3

The FREQ3 model has been computerized and is written in FORTRAN IV language for use on the CDC 6400 computer. The computer program consists of a main program that is essentially a calling program, 17 subroutines, and 1 function. A more detailed description of this model also is available (11).

The FORTRAN deck consists of approximately 2,000 statements. The computer time required for the FREQ3 program to process a 10-mile (16.1-km) section of congested freeway during a 2½-hour period (ten 15-min time slices) is approximately 4 sec. The computer program results have been calibrated with real-world data obtained from a number of sites including the northbound East Bayshore Freeway in the San Francisco Bay area. The output from the FREQ3 model includes speeds, densities, flows, and travel times for each combination of time slice and subsection; individual trip times and total travel times for each time slice; and total travel times and total travel distances for the entire freeway study section during the study period.

Table 1. PREFO program options.

Item	Option
Objective	1. Maximizing vehicle input rate 2. Maximizing vehicle miles of travel 3. Maximizing passenger input rate 4. Maximizing passenger-miles of travel
Formulation	1. Proportional diversion formulation 2. Short-trip diversion formulation
Diversion	1. No diversion 2. Total diversion
Number of O-D patterns	1. One O-D pattern for buses and automobiles 2. Two O-D patterns: 1 for buses, 1 for automobiles 3. Three O-D patterns: 1 for buses, 1 for automobiles with 1 passenger, and 1 for automobiles with 2 or more passengers 4. Four O-D patterns: 1 for buses, 1 for automobiles with 1 passenger, 1 for automobiles with 2 passengers, 1 for automobiles with 3 or more passengers 5. Five O-D patterns: 1 for buses, 1 for automobiles with 1 passenger, 1 for automobiles with 2 passengers, 1 for automobiles with 3 passengers, 1 for automobiles with 4 or more passengers 6. Six O-D patterns: 1 for buses, 1 for automobiles with 1 passenger, 1 for automobiles with 2 passengers, 1 for automobiles with 3 passengers, 1 for automobiles with 4 passengers, 1 for automobiles with 5 or more passengers
Main-line input fluctuation	1. Flow fluctuation considered 2. Flow fluctuation not considered

Composite Model, FREQ3CP

The integration of the FREQ3 and PREFO programs is called the FREQ3CP program. FREQ3CP consists of over 40 FORTRAN subroutines totaling about 3,500 cards. It has been implemented on the CDC 6400 and the IBM 360-65 computer systems and requires a real or virtual memory space of nearly 56,000 words if all subprograms are loaded together. The CDC 6400 at the University of California, Berkeley, computer center limits users to 40,978 words, so the model has been grouped into 3 segments according to its main simulation functions. The segment containing PREFO does not directly interface with the segment containing FREQ3 except through the main or root segment. This makes it possible to run the model in 40,000 words by "overlaying" the PREFO and FREQ3 segments. That is, while FREQ3 is being executed PREFO is retained in secondary (disk) storage, and vice versa. The root segment contains the program that governs the calling sequence of the other segments, and, in addition, it contains programs and data that are shared by the FREQ3 and PREFO segments.

The user has the choice of selecting any of the following options available in the FREQ3CP model:

1. Optimum control strategies (from PREFO submodel),
2. Freeway performance (from FREQ3 submodel), and
3. Optimum control strategies and freeway performance before and after control option (from PREFO and FREQ3 submodels).

APPLICATION OF FREQ3CP MODEL

The FREQ3CP model has been applied to 3 sites—the Santa Monica Freeway in Los Angeles, the East Bayshore Freeway in the San Francisco Bay area, and the Long Island Expressway on Long Island, New York. The purpose of applying the model was to demonstrate its great versatility and coincidentally to provide results that could be of use to the organizations that provided the data. This process demonstrated that the FREQ3CP model has 4 distinct purposes.

1. The FREQ3 option simulates normal freeway operations (no entry control).
2. The model provides optimum control strategies for regular ramp metering, that is, entry control on a vehicle basis.
3. The model provides optimum control strategies for priority-entry operations. Virtually any conceivable entry-control plan can be evaluated, including those that combine both vehicle control and priority control at different on-ramps in the freeway corridor. Another computer program, CPOD, is used to manipulate O-D tables if both types of entry control are to be used.
4. The FREQ3 simulates priority-lane operations on a freeway. Origin-destination tables are divided into priority vehicle and nonpirority vehicle tables through use of CPOD. Then the FREQ3 simulation is done separately for the reserved lanes and the unreserved lanes by using the appropriate set of O-D tables and correct capacities for the priority operations situation. If priority vehicles can enter the reserved lane at only 1 point, the PRIFRE model (10) should be used. Otherwise, the procedure described here is more appropriate to real-life situations.

It has been found that there are few situations involving priority operations that the FREQ3CP model cannot handle if the CPOD program is used in conjunction with the model. In fact, both entry-control and priority-lane operations can be evaluated virtually simultaneously by the model.

Site Description

The site chosen for the results to be presented here is the eastbound Santa Monica Freeway. This freeway is the busiest highway in the world; it carries up to a quarter of a million vehicles per day in both directions. It begins in Santa Monica and extends eastward about 13 miles (21 km) to an area near the Los Angeles CBD. The eastbound section investigated is about 9.5 miles (15.3 km) in length, extending from the interchange with the San Diego Freeway in west Los Angeles to the interchange with the Harbor Freeway near the Los Angeles CBD. There are 14 on-ramps and 14 off-ramps in this section of freeway. Under existing conditions congestion occurs daily on this section during the morning peak period. There are plans to control this freeway in the very near future.

Input Data

The input to the FREQ3CP model is of 3 types:

1. Freeway design parameters,
2. Freeway traffic-demand patterns (O-D tables), and
3. Linear programming objective and constraints and program options.

The freeway-design parameters and traffic-demand patterns were obtained from District 7, California Department of Transportation. From these data, the model was calibrated so that it accurately simulated existing conditions on the freeway. From the calibrated data and the computerized model, a series of analyses was performed to investigate both the short-term and long-term effects of priority-entry control and to compare the short-term effects to vehicle-entry control.

Short-Term Analysis

The first analysis involved a set of 4 computer runs, 1 for each of the 4 objective functions; constraints and program options were held constant. For this analysis the existing occupancy distribution and demand level were used because the short-term effects of priority control were of interest. The selected program options for this analysis are proportional diversion formulation, total diversion of all vehicles exceeding the optimal metering rate, and main-line input fluctuation with a 90 percent confidence interval and a 1.0 variance-to-mean ratio. Volumes were not allowed to exceed 0.99 of capacity.

The constraints were somewhat different for vehicle-entry control and priority-entry control. When control was on a vehicle basis, the maximum metering rate at 10 of the 14 ramps was 900 vehicles per hour (vph). At the 4 on-ramps with 2 lanes, it was possible to increase the maximum metering rate to 1,500 vph. The minimum metering rate at all on-ramps was 180 vph. These metering rates were the upper limit on metering capacity and the lowest possible rate to prevent excessive violation of the ramp signal. When control was on a priority basis, the metering rates had a different meaning. Previous analyses indicated that a priority cutoff level of 2 was the maximum that could be attained in this situation (and, probably, in most real-life situations). Thus a priority cutoff level of 2 was designated for all 14 ramps. This meant that all ramps would have priority entry, and that any vehicle with 2 or more occupants could enter the freeway without undergoing a queuing process. This would necessitate restriping or reconstructing all ramps that do not presently have 2 lanes, but this was not felt to be a serious constraint if priority-entry control were desired. The maximum and minimum metering rates for the nonpriority vehicles (those with a single occupant) were then set at 900 vph and 180 vph respectively. These metering rate constraints (for both vehicle control and priority control) were the limits that are feasible with metering, and a universal priority cutoff level of 2 is appropriate in practically any situation in which single-occupancy vehicles comprise at least two-thirds of all vehicles. In Los Angeles about 85 percent of all peak-period vehicles have only 1 occupant.

Levels of results will be given for the most critical time slice (7:15 to 7:30 a.m.) and the total peak period (6:30 to 9:30 a.m.).

Results for the 7:15 to 7:30 a.m. time slice are given in Tables 2 and 3, which give both the optimal metering rates for each ramp and the performance with regard to various measures of effectiveness for each of the 4 objective functions. The emphasis in this analysis is on short-term effects of priority control, that is, the situation 1 or 2 weeks after the concept is implemented. In the short term, entry control will eliminate congestion on the freeway, and priority vehicles will benefit from both a congestion-free freeway trip and their ability to bypass ramp queues of nonpriority vehicles.

Before control was exerted, the travel time from beginning to end of the freeway section was 17.8 min. After control, the travel time was reduced to 10.1 min. Although over 580 vehicles have been diverted from the freeway, vehicle miles (kilometers) of travel have increased by at least 4.6 percent. By comparing the various cases, it can be seen that, when vehicle miles (kilometers) or passenger-miles (kilometers) are

Table 2. Optimum metering rates for vehicle control and priority control, 7:15 to 7:30 a.m.

			Vehicle Control[a]				Priority Control[a]			
	Original Demand		Maximize Vehicle Input		Maximize Vehicle-Miles		Maximize Passenger Input		Maximize Passenger-Miles	
On-Ramp	vph	pph	vph	pph	vph	pph	vph	pph	vph	pph
Main line	7,200	8,460	7,200	8,460	7,200	8,460	7,200	8,460	7,200	8,460
1	952	1,119	900	1,058	900	1,058	952	1,120	952	1,120
2	676	794	676	794	676	794	676	795	676	795
3	504	592	180	212	504	592	256	344	504	593
4	504	592	504	592	504	592	504	593	504	593
5	300	352	180	212	300	353	225	278	300	353
6	1,300	1,527	1,173	1,379	1,191	1,399	925	1,154	1,062	1,290
7	532	625	180	212	180	212	260	353	260	353
8	1,060	1,245	1,060	1,246	598	703	1,059	1,246	599	785
9	1,320	1,551	688	809	683	802	691	924	685	917
10	1,000	1,175	891	1,047	888	1,044	772	948	768	944
11	800	940	185	218	180	212	300	441	300	441
12	440	517	440	517	440	517	440	517	440	517
13	540	634	540	635	540	635	540	635	540	635
14	400	470	400	470	400	470	400	470	400	470
Total	17,528	20,595	15,198	17,857	15,184	17,841	15,199	18,277	15,188	18,266

Note: 1 mile = 1.609 km.

[a]In a comparison of vehicle- and priority-control results, maximum metering rates differed.

Table 3. Measures of effectiveness for vehicle control and priority control.

Measure of Effectiveness	Original Demand	Vehicle Control: Maximize Vehicle Input	Vehicle Control: Maximize Vehicle-Miles	Priority Control: Maximize Passenger Input	Priority Control: Maximize Passenger-Miles
Passenger-miles	23,142	24,202	24,511	24,655	24,885
Vehicle-miles	19,696	20,597	20,860	20,642	20,874
Diverted demand, vph	—	2,330	2,344	2,329	2,340

Note: 1 mile = 1.609 km.

maximized, ramps farthest from the bottleneck are metered less restrictively than when input is maximized. But, to compensate, heavy control is imposed on ramps near the bottleneck, which occurs near ramp 11, and at least 40 percent of the demand at ramps 7, 9, 10, and 11 are denied entry to the freeway. When input is maximized, the severity of control is spread more evenly among the ramps along the freeway corridor although certain individual ramps fare no better. Because it is not desirable to restrict entry to only those freeway users from certain areas in the freeway corridor, input was maximized in subsequent analyses. In addition, a set of maximum and minimum metering rates that diverted no more than 100 vehicles from any ramp in any time period was prepared for use in a later analysis.

In the short term, the differences between priority-entry control and vehicle-entry control were not great, but neither were they insignificant. There was an increase of 2.4 percent in the number of persons able to use the freeway with priority control and small increases in both passenger-miles (kilometers) and vehicle miles (kilometers) of travel.

Priority control tends to treat ramps more equally than does vehicle control. Because both the priority-cutoff-level constraint and the minimum-metering-rate constraint must be satisfied, the optimum priority-control strategy allows greater input from the most restrictively controlled ramps than does a vehicle-control strategy.

Long-Term Analysis

The objective of priority-entry control is not merely to favor car pools. It also is intended to have long-term effects, namely, to induce those peak-period highway users who now travel alone to form car pools and thereby reduce vehicular demand. This will lead to a decrease in vehicle miles (kilometers) of travel, improved level of service, increased passenger capacity, and a reduction in air pollution from automotive sources. In conjunction with other techniques aimed at motivating increased car pooling, priority entry offers considerable promise in inducing commuters to shift to car-pool vehicles. In 2 actual cases of priority-entry operations in the Los Angeles area, new car pools were formed as a result of the implementation of a priority bypass lane at a freeway on-ramp. If major urban areas are to meet the ambient air quality standards for 1977 set by the U.S. Environmental Protection Agency (EPA), a reduction in vehicle miles (kilometers) of travel is essential, and in most cases this can best be accomplished by increased car pooling.

To determine what some likely consequences of priority-entry-stimulated car pooling would be, an analysis was made of the effect of various occupancy shifts on freeway corridor operations. An occupancy shift is defined as follows. All vehicles with 2 or more occupants were considered car-pool vehicles (the EPA definition of 3 or more seems unrealistic in many real-life situations). An x percent occupancy shift was defined as x percent of the persons in single-occupancy vehicles shifting into car-pool vehicles. The distribution of these persons among car-pool vehicles was in the same proportion as for the existing car-pool occupancy distribution. Thus passenger demand remained constant, but vehicle demand was reduced. Occupancy shifts of 3, 5, 10, 15, and 20 percent were analyzed. Originally 85 percent of the vehicles had 1 occupant, and this was successively reduced to 75 percent (for the 20 percent occupancy shift). FREQ3CP was used to analyze all such cases, and the results of these analyses are given in Tables 4 and 5.

Table 4 illustrates the effects of the various occupancy shifts on certain measures of performance. To determine the effect of occupancy shifts on reductions in vehicle miles (kilometers) of travel for the freeway corridor (a necessity because some demand is diverted to arterial streets) it was found that the average trip length of diverted vehicles was about 2.75 miles (4.4 km). This was multiplied by the number of diverted vehicles and added to the freeway vehicle miles (kilometers) of travel. An average speed of 20 mph (32.2 km/h) was assumed for travel on the arterial streets, which permitted the calculation of vehicle hours expended by the diverted vehicles. For the base case with priority-entry control, the diverted demand represented 3.5 percent of the total demand and was spread over 9 of the 12 time slices. Conversations with Los Angeles officials confirmed that the surface street system would have little difficulty in absorbing these additional vehicles.

It has been concluded that a 10 percent occupancy shift is attainable if car pooling is aggressively pursued in Los Angeles (2). A 3 percent occupancy shift is the minimum likely (2), and a 20 percent shift seems to be the upper limit unless coercive policies are adopted. As Table 4 indicates, a 10 percent occupancy shift reduces vehicle miles (kilometers) of travel by 4.6 percent over the present situation. It also results in a 41 percent decrease in vehicle hours expended by the present demand. A 20 percent occupancy shift would reduce vehicle miles (kilometers) of travel by 9.1 percent and vehicle hours by 44 percent compared to the present situation. The occupancy shifts also increase the productivity of the freeway compared to the base case with priority control. A 10 percent shift increases freeway passenger-miles (kilometers) by 0.5 percent and reduces necessary diversion by 34 percent compared to priority control with no occupancy shift. A 20 percent shift increases freeway passenger-miles (kilometers) by 1.0 percent and reduces diversion by 67 percent. Priority-entry control promises to provide substantial travel-time savings to peak-period commuters, and, if occupancy shifts occur, they will reduce vehicle miles (kilometers) of travel by amounts that could be considered significant in Los Angeles. And, as given in Table 2, priority-entry control also makes more effective use of the freeway in terms of both people and vehicles.

What will be the motivation for these occupancy shifts? Table 5 indicates that travel-time savings could be a very important motivation. Previously, we have made the unrealistic assumption that all traffic in excess of that permitted by the optimum metering rates diverted to the surface streets. In actual experience, some vehicles do not divert; they wait in the ramp queue before gaining entrance to the freeway. Ramp delays of 5 min or more are common in Los Angeles. In Table 5, we assumed that 60 percent of the excess demand at the Washington Boulevard on-ramp would divert and that the remainder would queue up. The travel times for the nonpriority vehicles reflect this ramp queue

Table 4. Measures of effectiveness for different levels of shifts, 6:30 to 9:30 a.m.

Case	Freeway (passenger-miles)	Corridor (vehicle-miles)	Corridor (vehicle hours)	Diverted (vehicle demand)
Base, normal operations	266,902	227,151	6,837	–
Base, vehicle control	262,725	227,151	4,335	1,572[a]
Base, priority control	263,198	227,151	4,357	1,621[a]
3 percent shift, priority control	263,490	224,016	4,270	1,517
5 percent shift, priority control	263,842	221,949	4,198	1,360
10 percent shift, priority control	264,506	216,747	4,036	1,065
15 percent shift, priority control	265,058	211,546	3,916	807
20 percent shift, priority control	265,712	206,367	3,808	529

Note: 1 mile = 1.609 km.

[a]These rates are different because of difference in maximum metering rate.

Table 5. Travel time, Washington Boulevard to Harbor Freeway, 7:30 to 7:45 a.m.

Case	Priority Vehicles (min)	Nonpriority Vehicles, Assuming 60 Percent Diversion (min)	Savings for Priority Vehicles (min)
Priority Control, Optimal Metering Rates			
Base	5.99	11.38	5.39
3 percent shift	5.97	10.05	4.08
10 percent shift	5.90	7.38	1.48
20 percent shift	5.86	5.86	0
Priority Control, Equalized Metering Rates			
Base	5.92	12.58	6.66
3 percent shift	5.90	11.68	5.78
10 percent shift	5.87	9.75	3.88
20 percent shift	5.82	7.84	2.02

Note: Travel time before control: 12.25 min. Assumed street travel time: 15 min.

delay, which is the travel-time savings for the priority vehicles. In spite of the ramp delays, total travel time is less than that likely by surface street. Two cases were analyzed: One used optimal metering rates for the maximum and minimum metering constraints previously discussed (Case A); the other, diverted demand (Case B), was spread evenly among the various ramps. The rates used for Case B were probably closer to those that would be used in the field than were those for Case A for practical considerations. In the short term (base case and 3 percent shift) there were substantial travel-time savings for priority vehicles in either case. Nonpriority vehicles experienced travel times 70 to 110 percent greater than those for priority vehicles, which should be a major inducement for the formation of car pools. Even with a 10 percent shift, priority vehicles save a significant amount of time in Case B. Only at a 20 percent shift is the travel-time motivation rather insignificant. And even if all present freeway demand were served as a result of occupancy shifts, the latent demand on the parallel arterial routes probably would divert to the freeway. This would again cause ramp queues for nonpriority vehicles, and continue the incentives for occupancy shifts. There is undoubtedly a point at which the freeway could serve all present likely demand (both manifest and latent), but this would be for occupancy shifts considerably higher than those considered here. (A 45 percent occupancy shift would be necessary to increase auto occupancy to 1.5 from the present 1.18.) Thus there will probably always be travel-time incentives to car pool if priority entry is implemented. At the same time, the occupancy shifts also will benefit those who must still drive alone by creating shorter waiting times at the freeway ramps.

Conclusion

The FREQ3CP model has demonstrated its versatility and usefulness in our analyses. Even more significantly, it has been shown that the concept of priority-entry control can achieve several important objectives. In the short term, priority-entry control will increase person use of a freeway, eliminate freeway congestion, reduce vehicle hours expended on the freeway, and result in significant travel-time savings for car-pool vehicles. In the long term, if priority-entry control is implemented in conjunction with other techniques to motivate increased car pooling, reductions in vehicular demand will occur which, in turn, will decrease automotive emissions, increase level of service, and increase passenger capacity. Thus priority-entry control promises favorable effects in terms of both improvements in freeway traffic operations and reductions in air pollution.

SUMMARY

This paper proposed a new control technique for urban freeways, priority-entry control, which promises to provide immediate benefits, to modify future demands, and to provide even greater long-term benefits. The immediate benefits are an increase in passenger capacity and a reduction in passenger travel time. The long-term benefits include increased vehicle occupancy by encouraging occupants in low-occupancy vehicles to change to higher occupancy vehicles. Such changes will significantly reduce energy consumption and air pollution per passenger-mile (kilometer) of travel.

The solution of this priority-entry control problem is formulated as a linear programming problem that is very flexible and permits the selection and use of a wide variety of objectives and constraints. By the use of an upper-bounding method, the solution to the formulation is made very efficient.

The linear programming formulation was computerized and integrated with a previously developed and tested traffic performance simulation model. The integrated computerized program is called FREQ3CP and can be applied to freeway sections up to 10 miles (16.1 km) in length and for multitime slices.

The integrated computerized program was applied to 3 typical, heavily congested urban freeways, and a number of investigations were undertaken to demonstrate the applicability and flexibility of the methodology. The benefits of priority-entry control over normal ramp control strategies were demonstrated.

Although considerable progress has been made in developing a methodology for priority-entry control strategies, there are, nevertheless, ways for improving and extending the methodology. The 2 most important areas for future research are traffic-demand transfer between time slices and traffic-demand transfer between alternative routes during periods of priority-entry control. Essentially this methodology requires diversion and assignment submodels that can operate on a freeway-corridor basis. In addition, a modal-split submodel is needed to estimate vehicle occupancy demand as a function of the priority-entry-control strategy.

REFERENCES

1. G. B. Goodel. Preferential Access for Multi-Occupant Vehicles at Metered On-Ramps. Traffic Engineering, Vol. 44, No. 12, Sept. 1974.
2. A Study of Techniques to Increase Commuter Vehicle Occupancy on the Hollywood Freeway. Alan M. Voorhees and Associates Inc. Final Rept., Nov. 1973.
3. J. A. Wattleworth and D. S. Berry. Peak-Period Control of a Freeway System—Some Theoretical Investigations. Highway Research Record 89, 1965, pp. 1-25.
4. M. E. Goolsby, E. Merrell, and W. R. McCasland. Freeway Operations on the Gulf Freeway Ramp Control System. Texas Transportation Institute, Texas A&M Univ., Research Rept. 24-25, 1969.
5. K. A. Brewer, J. H. Buhr, D. R. Drew, and C. J. Messer. Ramp Capacity and Service Volume as Related to Freeway Control. Highway Research Record 279, 1969, pp. 70-86.
6. J. J. Wang and A. D. May. Freeway Operations Study Phase III, Analysis of Freeway On-Ramp Control Strategies. Rept. No. 73-3, June 1973.
7. K. Ovaici and A. D. May. Freeway Priority Entry Control to Favor Multi-Passenger Vehicles. 6th International Symposium on Transportation and Traffic Theory, Sydney, Australia, Aug. 1974.
8. Highway Capacity Manual. HRB Spec. Rept. 87, 1965.
9. K. Ovaici, R. F. Teal, J. K. Ray, and A. D. May. Priority Entry Control of Freeway Ramps for High Occupancy Vehicles. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, Dec. 1974.
10. D. R. Minister, L. P. Lew, K. Ovaici, and A. D. May. A Computer Simulation Model for Evaluating Priority Operations on Freeways. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, June 1972.
11. W. A. Stock, R. C. Blankenhorn, and A. D. May. The FREQ3 Model. In Freeway Operations Study—Phase III, Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, June 1973.

SPONSORSHIP OF THIS RECORD

GROUP 3–OPERATION AND MAINTENANCE OF TRANSPORTATION FACILITIES
Lloyd G. Byrd, Byrd, Tallamy, MacDonald, and Lewis, chairman

Committee on Freeway Operations
Robert S. Foote, The Port Authority of New York and New Jersey, chairman
Donald E. Orne, Michigan Department of State Highways and Transportation, secretary
Patrick J. Athol, John L. Barker, Elmer N. Burns, J. R. Doughty, Conrad L. Dudek, A. C. Estep, Paul H. Fowler, Joseph W. Hess, E. Vinson Hoddinott, Hugo O. Liem, Jr., Alger F. Malo, Adolf D. May, Jr., William R. McCasland, Joseph M. McDermott, Stuart F. Millendorf, Ann Muzyka, Eugene F. Reilly, Robert A. Reiss, D. H. Roper, Robert E. Shields, T. Darcy Sullivan, A. Taragin, Joseph Treiterer, Joseph A. Wattleworth, Robert H. Whitson, H. Nathan Yagoda

K. B. Johns, Transportation Research Board staff

The organizational units and the chairmen and members are as of December 31, 1974.