NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM REPORT **28**

SURVEILLANCE METHODS AND WAYS AND MEANS OF COMMUNICATING WITH DRIVERS

HIGHWAY RESEARCH BOARD
NATIONAL RESEARCH COUNCIL
NATIONAL ACADEMY OF SCIENCES—NATIONAL ACADEMY OF ENGINEERING

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM REPORT **28**

SURVEILLANCE METHODS AND WAYS AND MEANS OF COMMUNICATING WITH DRIVERS

MORTON I. WEINBERG, NORMAN J. DELEYS, AND
RICHARD F. SCHNEEBERGER
CORNELL AERONAUTICAL LABORATORY
BUFFALO, NEW YORK

RESEARCH SPONSORED BY THE AMERICAN ASSOCIATION
OF STATE HIGHWAY OFFICIALS IN COOPERATION
WITH THE BUREAU OF PUBLIC ROADS

SUBJECT CLASSIFICATION:
- TRAFFIC CONTROL AND OPERATIONS
- TRAFFIC FLOW
- TRAFFIC MEASUREMENTS

HIGHWAY RESEARCH BOARD
DIVISION OF ENGINEERING NATIONAL RESEARCH COUNCIL
NATIONAL ACADEMY OF SCIENCES—NATIONAL ACADEMY OF ENGINEERING 1966

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

Systematic, well-designed research provides the most effective approach to the solution of many problems facing highway administrators and engineers. Often, highway problems are of local interest and can best be studied by highway departments individually or in cooperation with their state universities and others. However, the accelerating growth of highway transportation develops increasingly complex problems of wide interest to highway authorities. These problems are best studied through a coordinated program of cooperative research.

In recognition of these needs, the highway administrators of the American Association of State Highway Officials initiated in 1962 an objective national highway research program employing modern scientific techniques. This program is supported on a continuing basis by funds from participating member states of the Association and it receives the full cooperation and support of the Bureau of Public Roads, United States Department of Commerce.

The Highway Research Board of the National Academy of Sciences-National Research Council was requested by the Association to administer the research program because of the Board's recognized objectivity and understanding of modern research practices. The Board is uniquely suited for this purpose as: it maintains an extensive committee structure from which authorities on any highway transportation subject may be drawn; it possesses avenues of communications and cooperation with federal, state, and local governmental agencies, universities, and industry; its relationship to its parent organization, the National Academy of Sciences, a private, nonprofit institution, is an insurance of objectivity; it maintains a full-time research correlation staff of specialists in highway transportation matters to bring the findings of research directly to those who are in a position to use them.

The program is developed on the basis of research needs identified by chief administrators of the highway departments and by committees of AASHO. Each year, specific areas of research needs to be included in the program are proposed to the Academy and the Board by the American Association of State Highway Officials. Research projects to fulfill these needs are defined by the Board, and qualified research agencies are selected from those that have submitted proposals. Administration and surveillance of research contracts are responsibilities of the Academy and its Highway Research Board.

The needs for highway research are many, and the National Cooperative Highway Research Program can make significant contributions to the solution of highway transportation problems of mutual concern to many responsible groups. The program, however, is intended to complement rather than to substitute for or duplicate other highway research programs.

This report is one of a series of reports issued from a continuing research program conducted under a three-way agreement entered into in June 1962 by and among the National Academy of Sciences-National Research Council, the American Association of State Highway Officials, and the U. S. Bureau of Public Roads. Individual fiscal agreements are executed annually by the Academy-Research Council, the Bureau of Public Roads, and participating state highway departments, members of the American Association of State Highway Officials.

This report was prepared by the contracting research agency. It has been reviewed by the appropriate Advisory Panel for clarity, documentation, and fulfillment of the contract. It has been accepted by the Highway Research Board and published in the interest of an effectual dissemination of findings and their application in the formulation of policies, procedures, and practices in the subject problem area.

NCHRP Project 3-2 FY '63
NAS-NRC Publication 1473
Library of Congress Catalog Card Number: 66-62658

FOREWORD

By Staff

Highway Research Board

This final report will be of interest to traffic engineers and to highway officials responsible for the operation of high-volume freeways and congested urban streets. It presents a method of predicting freeway travel times when the freeway is operating at a very low level of service. Also included is a detailed study of the effectiveness of the airborne observer in traffic control. It is believed that this research effort is one of the most comprehensive studies to date of modern surveillance methods and ways and means of communicating with the driver. The research results contained herein will aid engineers in the design, development, and modernization of improved traffic surveillance systems. Although this report describes the findings of this project, a more complete documentation of the research conducted during the first year is presented in NCHRP Report No. 9.

This report stems from the NCHRP project entitled "Surveillance Methods and Ways and Means of Communicating with Drivers." Only the findings of this project that pertain to freeway surveillance and the airborne observer in traffic control are reported herein. The final report for the studies involving a digital-computer-controlled traffic signal system for a small city is presented in NCHRP Report No. 29.

Although the initial freeway concept was developed on the premise that traffic would operate unhindered by traffic controls, it is now apparent that properly designed control systems will improve the freeways' ability to service the ever increasing traffic demand. For this reason research was initiated to consider closed-loop control systems. The surveillance schemes utilized must be capable of diverting the traffic attempting to use the congested facility onto alternate routes. Three types of closed-loop surveillance systems have been studied in depth.

The first involves studies on the John C. Lodge Freeway in Detroit. In this system freeway stream flow parameters are measured, and a digital computer, programmed in real time, analyzes the incoming field data, predicts future travel times, and automatically activates controls on the freeway and the surrounding street system. This study is limited, inasmuch as no attempt was made to actually practice the control technique; however, this research shows that sufficiently accurate travel time predictions can be made to implement a practical freeway early warning system that will enable motorists to avoid the congested freeway. The research presents for the first time the development of the mathematical logic necessary to predict travel times on an urban freeway in the event that a mishap causes a breakdown in the traffic flow.

The second surveillance system researched concerns the effectiveness of using an airborne observer to control traffic. Radio broadcasts inform the motorists of the most desirable traffic routing and cost studies are made utilizing a light airplane and a helicopter. This is one of the first studies providing detailed information pertaining to the effectiveness of airborne surveillance. It is shown that the savings in motor vehicle operating costs far exceed the operating costs of this system. Furthermore, it is concluded that a balloon-borne closed-circuit television surveillance scheme is a logical extension of the airborne observer concept and a preliminary design of this system is presented.

The third type of closed-loop surveillance system studied deals with the digital-computer-controlled traffic signal system and its design. The results of these studies are presented in NCHRP Report No. 29.

The report notes that travel times can be predicted that correlate quite closely with observed travel times. As a result several fruitful research areas have become evident. It is suggested that future research should involve a test of the experimental logic with electronic computer equipment in real time on a freeway. This should also involve the development and implementation of an experimental control system and the evaluation of its effects on traffic.

CONTENTS

1 SUMMARY

2 CHAPTER ONE General Discussion
Surveillance on John C. Lodge Freeway
The Airborne Observer in Traffic Control

3 CHAPTER TWO Prediction of Travel Time on an Urban Freeway
Objectives
Approach
Implementation of Technical Approach
Data Processing
Correlation of Computed with Observed Travel Times
Discussion of Initial Results
Discussion of Final Results

13 CHAPTER THREE The Airborne Observer in Traffic Control
Project SCOUT
Light Airplane Operation
Combined Helicopter-Airplane Operation
Method of Analysis of an Incident
Analysis of a Traffic Tie-Up
Travel Time Data from Project SCOUT
Reconstruction of the Traffic Situation
Dissipation of the Queue
Calculation of Delay Time in Queue
Possible Situation—No Trafficopter
Evaluation of Trafficopter Effectiveness
Results of Cost-Effectiveness Study

21 CHAPTER FOUR Preliminary Investigation of an Airborne Television System for Traffic Surveillance
Description of the Concept
Some Design Considerations
Preliminary Design of the System
Preliminary Estimate of Weights
Preliminary Estimate of Costs

25 REFERENCES

25 APPENDIX A Surveillance Logic
1. Mathematical Logic for Computing Travel Time on an Urban Freeway
Highway Capacity Criterion
Mathematical Derivation of the Computational Logic
Discussion of Measured Traffic Characteristics
Criteria for Prediction Computation
Formation of the Queue
Dissipation of the Queue
Computation of Average Speed
Outline of the Computational Procedure

2. Travel Time as a Measure of Road Network Efficiency and the Acquisition of Efficiency Elements
Efficiency
Elementary Efficiencies
Elementary Indices
Extension to All *M* Exits
Measurement of Efficiency Elements
Concluding Comments

3. Implications of a Linear Speed-Density Relationship
Notation
Linear Assumption
Volume
Vehicle Spacing
Transit Time
Time Headway

4. Linear Theory of Traffic Flow and Replacement by a Valid Transmission Function
The Problem
Criticism of Linear Model
Two Requirements for a Valid Theory
Transmission Function
Tandem Requirement Satisfied
Discussion of Transmission Function

5. The Transmission Function as a Model for Study of Traffic Networks
Present Approach
Traffic Network Model
Entering and Leaving Functions
Discussion of Fundamental Equation
Transmission Function Concept
Network Interpretation of Transmission Function Concept
Probability Interpretation of Transmission Function
Practical Application

48 APPENDIX B Freeway Surveillance Studies

1. Applications of Real Highway Data to Traffic Congestion Computational Logic
First Application
Second Application

2. Automatic Detection of Stopped Vehicles by Electronic Means

3. Acoustic and Radio-Frequency Presence Detector Limitations
Presence Detectors
Sensitivity of Speed Determination

FIGURES

4 Figure 1. Typical speed-volume relationship.
5 Figure 2. Typical speed-volume curve.
6 Figure 3. Instrumented portion of John C. Lodge Freeway, Detroit.
8 Figure 4. Traffic behavior mapped on volume-density curve.
10 Figure 5. Sample observed travel time data sheet.
11 Figure 6. Sample predicted travel time print-out.
12 Figure 7. Presentation of results.
12 Figure 8. Volume-density curves for dry and rainy weather.
13 Figure 9. Presentation of additional results.
17 Figure 10. Typical data cards from Project SCOUT.
18 Figure 11. General plan of Buffalo Skyway route.
20 Figure 12. Comparison of time saving by use of Trafficopter with normal traffic tie-up.
22 Figure 13. Interchange at junction of Lodge and Ford Freeways, Detroit.
24 Figure 14. General view of proposed balloon installation.
24 Figure 15. Camera support platform and tilt drive assembly.
26 Figure A-1. Typical speed-volume relationship.
29 Figure A-2. Geometry of bottleneck occurrence.
29 Figure A-3.
33 Figure A-4. Typical subnetworks for traffic flow analysis.
33 Figure A-5. Efficiency at exit i in terms of elementary efficiencies.
33 Figure A-6. Index at exit i in terms of elementary indices.
34 Figure A-7. Parallel representation of efficiency (upper) and series representation of index (lower).
36 Figure A-8. Speed vs density, on linear assumption.
36 Figure A-9.
37 Figure A-10. Relative speed and relative density vs relative spacing.
37 Figure A-11.
38 Figure A-12. Transit time and spacing vs relative speed and relative density.
39 Figure A-13. Road traffic network.
39 Figure A-14. Entering and leaving functions on assumption $L(t) \propto Q(t)$.
40 Figure A-15.
40 Figure A-16. Input, output, and transfer functions.
42 Figure A-17. Examples of traffic networks.
43 Figure A-18. Application of $e(t)$ and $l(t)$ to more basic networks.
44 Figure A-19. Examples of law of conservation of vehicles under indicated assumptions.
44 Figure A-20. Network depicted as cause and effect.
45 Figure A-21. Elementary section of highway.
45 Figure A-22. Ideal performance of traffic network.
45 Figure A-23. Electrical network analogue of ideal performance.
47 Figure A-24.
47 Figure A-25. Electrical network analogue of probable performance.
50 Figure B-1. Distribution of lane volumes on Congress Street Expressway, Chicago, at East Avenue, 3:00 to 4:00 p.m.
51 Figure B-2. Volume-density curve, showing constant volume and constant speed relationships.
53 Figure B-3. Hourly volume variations on westbound Congress Street Expressway at East Avenue, October 4-24, 1962.

FIGURES

53 Figure B-4. Volume-density curve, showing relationships involved.

56 Figure B-5. Volume vs average density.

57 Figure B-6.

58 Figure B-7.

58 Figure B-8.

58 Figure B-9.

60 Figure B-10. Speed and density variation, John Lodge Freeway.

61 Figure B-11. Speed vs density, at 1-min intervals, John Lodge Freeway, October 8, 1963, 3:00 to 6:00 p.m.

63 Figure B-12. Signal process timing for over-the-roadway presence detector.

63 Figure B-13. Occupancy relationship under two detection pulse schemes.

64 Figure B-14. Maximum and minimum limits for pulse echoes.

65 Figure B-15. Computed vs actual speeds.

65 Figure B-16. Effect of beam width on occupancy measurement.

65 Figure B-17. How beam width affects echo of detector pulse.

ACKNOWLEDGMENTS

Morton I. Weinberg, Head, Transportation Systems Section, had primary responsibility for the technical direction and administration of the project reported in both this report and NCHRP Report 29. Other Cornell Aeronautical Laboratory staff personnel working on the portions of the project reported herein were as follows:

Norman J. DeLeys, Research Mechanical Engineer, who performed the preliminary analysis for the design of the airborne television system for traffic surveillance; Richard F. Schneeburger, Head, Intersciences Applications Section, who assisted in resolving equipment characteristics and applicability of basic electronic techniques; and John P. Persico, Associate Mathematician.

Appreciation is expressed to the many individuals and companies who provided the project team with valuable assistance.

Special acknowledgment is given to the following:

Edward F. Gervais, Project Manager, National Proving Ground, and Conrad L. Dudek, Highway Traffic Engineer, both of the Michigan State Highway Department, for their cooperation in the study conducted on the John C. Lodge Freeway.

John R. Sharpe, Newscaster, and David F. Leopold, Manager, both of Radio Station WEBR; Cyril B. King, Editor of the *Buffalo* (New York) *Courier-Express;* and William H. Schneider, Commissioner of Police, Buffalo, N. Y., for their contributions in the evaluation of the airborne observer.

Vernon H. Stone, Manager, Inflatable Systems Department of Litton Industries, and Howard Coyne, Cleveland Regional Office, Federal Aviation Agency, for supplying design data for use in the preliminary design of the balloon-borne, closed-circuit television system.

SURVEILLANCE METHODS AND WAYS AND MEANS OF COMMUNICATING WITH DRIVERS

SUMMARY

This study is concerned with researches on traffic surveillance systems for urban freeways. Investigations have been performed on (a) a method for predicting travel time on an urban freeway in the event of a traffic tie-up, and (b) the effectiveness of an airborne observer to reroute traffic in the event of a traffic tie-up.

Chapter One of the report establishes continuity between the first and present phases of this project, describing briefly the work done during the first phase that is used as a preamble to the two continuing tasks noted in the previous paragraph.

Chapter Two describes the development of a mathematical logic to predict travel time on an urban freeway in the event that a mishap causes a breakdown in traffic flow. The calculation of the conditions that indicate a tie-up and the resulting travel times are to be performed in real time on a high-speed digital computer. Data on traffic behavior are transmitted to the computer from volume-occupancy sensors located over the lanes of the freeway. The spacing of sensor stations along the length of the freeway is intended to be consistent with the data sampling period and the expected mean speed of the vehicles. The computer is programmed to estimate the time required for vehicles presently located at various sensor stations to pass through the estimated position of the bottleneck.

The object of this research was to devise, for the John C. Lodge Freeway in Detroit, an "early warning" system that could provide information on which to base advisories to motorists to avoid the freeway and seek alternate routes.

The computational procedure was programmed for solution on the computing equipment at the Cornell Aeronautical Laboratory. Data from 15 sensors at 5 stations along the freeway were recorded at 60-sec intervals during periods of peak traffic flow. During these same periods visual observations of congested conditions on the Freeway were made from the television monitors that view 3.2 miles of its length. When the computing equipment installed at the Freeway Control Center indicated that tie-up conditions existed, the observers timed the progress of vehicles through the 5 sensor stations. From these observations it was found that the observed travel times could be predicted. The predicted travel times varied from 5 percent less to 70 percent more than the observed travel times. A histogram of the numbers of events, classified in 10 percent intervals, shows no classical shape of statistical distribution, although the average value of the error was close to zero.

Chapter Three reports on the program to evaluate the effectiveness of an airborne observer to minimize delay by issuing alternate route advisories to motorists in the event that an incident threatens to tie up traffic. Communication was by means of commercial radio. Two crude analyses were performed during the initial phase of the project that seemed to indicate that a helicopter-observer system, broadcasting through a single radio station, could justify its cost of about $35,000 per year. It was concluded that a much broader data base was needed in order to increase the accuracy of the analysis of the traffic situation.

Project SCOUT enlisted the aid of nearly 5,800 motorists who, in addition to

the police and sheriff's departments and fire department, were alerted to report on pertinent happenings when radio station WEBR's "Trafficopter" advised them of traffic difficulties in the Buffalo area and recommended alternate routes of travel. Over a 6-month period there were 13 incidents, for which 120 reports were made. Nine of the incidents were analyzed, resulting in an estimate of an average saving of $465 in delay time for the vehicles alone, based on a cost of $1.25 per vehicle-hour.

A light airplane was used in place of the helicopter for 24 1-hr flights. It was determined that the airplane probably could be used for 80 percent of the flights, the helicopter being needed for flying in poor weather. A 500-hr-per-year patrol using both aircraft would cost about $15,000, plus an initial outlay of about $2,000 for two VHF two-way radios.

A long-term log of Trafficopter broadcasts showed that about 100 incidents per year could be expected, for an annual saving in delay time estimated at $48,000 for the vehicles alone.

An extension of the airborne observer concept is described in Chapter Four. Preliminary design has been accomplished for a surveillance scheme in which a closed-circuit television camera is borne aloft by a tethered balloon. A three-wire anchorage is to provide position and orientation stabilization in winds up to 20 mph. Some of the design limitations are discussed. The cost for the 40-ft diameter balloon system, to be raised to an altitude of 400 ft, is about $15,000.

CHAPTER ONE

GENERAL DISCUSSION

The foreword of NCHRP Report No. 9 (*1*) closed with ". . . The next phase will include a study of freeway surveillance, a study of digital-computer-controlled traffic signal networks, further study of the requirements of an airborne-observer surveillance and control system, and a study of driver communication by aural and visual messages." Of the four tasks named, only the last has not been pursued. The results of the studies involving the digital-computer-controlled traffic signal network are presented in NCHRP Report No. 29 (*2*).

A brief review of the previous work done on each of the tasks that were continued is given in the present report. However, for maximum continuity of documentation the reader should peruse NCHRP Report No. 9 (*1*) as a prelude to this report.

SURVEILLANCE ON JOHN C. LODGE FREEWAY

The portion of the project dealing with surveillance on the John C. Lodge Freeway was the continuation of an experimental study conducted in an effort to develop a real-time density prediction logic that would give advance warning of impending "congestion" on a freeway. The computational logic depended on traffic volume counts being sensed by detectors located over the lanes of the freeway at appropriate intervals. Data to exercise the logic were furnished by the Expressway Surveillance Project for the Congress Street Expressway in Chicago and by the John C. Lodge Freeway Control Center in Detroit. Although some indication of promise for the density prediction concept was revealed, it was too demanding of the instrumentation available to warrant further pursuit. However, it was concluded that a computational logic could be written that would be compatible with currently available traffic sensing devices.

Travel time prediction logic has been developed. It is intended to be supplied with data from presence detectors, one for each lane in the freeway at each traffic sensing station. For a validation of the mathematical model and the computational logic, detectors were installed on the John C. Lodge Freeway in Detroit and a research effort was integrated into the program of the National Proving Grounds for Freeway Surveillance, Control, and Electronic Traffic Aids (NPG). A complete description of the task is the subject of Chapter Two of this report.

THE AIRBORNE OBSERVER IN TRAFFIC CONTROL

The first study, published in NCHRP Report No. 9 (*1*), was an attempt to evaluate the cost-effectiveness of an airborne observer in directing traffic in emergency situations.

The observer was the radio station WEBR "Trafficopter," operating in the Buffalo, N. Y., area, and using a helicopter as the aerial platform. Inasmuch as this kind of traffic advisory service is becoming more widely used, research was conducted to determine the benefits realized from the installation of an airborne-observer service supported by public funds.

The first study led to the conclusion that adequately accurate analyses of traffic behavior, responding to the airborne observer, would require a significant broadening of the data base. First, attempts to quantify travel times, numbers of vehicles experiencing delay, numbers of vehicles following alternate routes, formation and dissipation of queues, etc., depended on spotty information received from the Trafficopter, police and sheriff's departments, fire department, and, rarely, a motorist. Therefore, in the second study, presented in Chapter Three, the data base was extended by inviting all of the motorists in the Buffalo area to become "deputy researchers."

The advantages of an area (two-dimensional) view of traffic have been known for some time and this research points out the value of the airborne human observer. Photography is presently being used throughout the nation to obtain traffic data, and this technique is widely accepted. However, photography is an "after the fact" communication process, and thus is not appropriate for "real time" traffic analysis and control.

Television cameras mounted on fixed structures have obtained video presentations that can provide information for real-time control of traffic. Research conducted on the John C. Lodge Freeway has shown the value of closed-circuit video. It is also possible to mount a television camera in an airplane or helicopter and to telemeter the video to ground observers. This would provide for excellent coverage (properly located buildings or towers are not always available) and at great speed, but it is expensive for long-term operation and includes some hazard to life.

Considering the technical, operational, and economic problems involved, this research concludes that a balloon-borne closed-circuit television system should be considered as a new technique for area surveillance. A preliminary design for this type of system is presented in Chapter Four.

CHAPTER TWO

PREDICTION OF TRAVEL TIME ON AN URBAN FREEWAY

Because freeways are able to carry large volumes of traffic during the peak periods, the total highway network that they serve is heavily dependent on their continued operation at high performance levels. Programs have been instituted to study means for insuring that high levels of service be maintained on freeways. Typical are the researches being accomplished in Chicago, Detroit, and Houston, where real-time surveillance and control of traffic is being studied on operational facilities. Basic work has been done in traffic flow, correlating theory with observation. The effects of lane closure signals to give advance warning of blockages, ramp controls, speed controls, etc., have been studied, all to the end that maximum speed and volume, consistent with safe operation, might be maintained. Surveillance has been accomplished by means of over-the-road detectors, aerial photography, and closed-circuit television.

But the research has even larger goals in view. Not only are the freeways to operate efficiently, but the combination of the freeways and the street complex into which they are integrated are to be controlled to give optimum service.

The end "product" desired is a full-time, real-time, surveillance and control system using automated data acquisition and processing. Traffic control "decisions" would be made by high-speed digital computers into which appropriate logics for control would have been programmed. Communication of control would be made to the driver by visual (signs and signals) and aural (radio) means as appropriate. Optimum control is defined as that assignment of right-of-way in the entire complex that permits all vehicles to move to their destinations at maximum average speed (minimum aggregate delay) within two constraints. One constraint is the provision for pedestrian crossings, the other is that no individual vehicle deliberately shall be denied passage in excess of some arbitrarily chosen period of time.

The research work in progress throughout the nation may be considered as building blocks, each project contributing to the final goal. After assessing the programs being pursued on the several freeways and in network control in Toronto, it was concluded that a required adjunct to a control system would be a "warning" scheme that provides advance notification of the need to re-route traffic. The need for such a scheme is greatest for the high-volume freeway, where an unforeseen breakdown in traffic flow has far-reaching effects on many vehicles. Accordingly, this project, working in cooperation with personnel of the National Proving Ground, has devised a method for predicting travel time on the John C. Lodge Freeway in Detroit in the event that an unexepected breakdown in flow occurs. The development of the prediction logic and the method for its validation are the subject of this chapter.

OBJECTIVES

The objective is to determine, in the event of an incident, the reduction of service level that will occur on the freeway, by a forecast of travel time, and to signify when the forecast indicates that the motorists would be better served by using the surface street system. The general task may be stated as:

1. To define that travel time along the freeway that can be equalled by travel along alternate routes (surface arterials).
2. To develop a traffic surveillance procedure and computational logic that will monitor and predict traffic behavior in order to implement controls to divert traffic from the freeway when the travel time threatens to exceed that in item 1.

The present part of the task is concerned with the second item.

APPROACH

The average speed-headway relationship mandated by drivers leads to a speed-volume curve that is roughly parabolic in shape (Fig. 1). Furthermore, in a one-dimensional infinitely long stream flow maximum volume operation of a highway serves the "greatest good for the greatest number." Therefore, for any particular highway segment, environmental condition, and driver population it is desirable to keep the flow at maximum volume, assuming that a "saturation" demand exists.

During peak demand periods it is desirable that the speed of travel on the freeway be maintained at that value consistent with maximum volume. This speed is designated as "critical speed." Detection and computing equipment can monitor the traffic behavior in real time, and can indicate current average speed and volume over preselected time periods. Speed and volume values can be compared with values derived statistically within a recent past period to determine if the road is operating at conditions conducive to maximum volume.

If a reduction in volume is sensed and computed, it is necessary to determine whether or not it was caused by the free will of the drivers. If it was not, controls must be activated. If average speed has increased, demand has fallen off and controls are not necessary. If the speed has remained at critical value, this too indicates a reduction in demand. However, if a reduction in speed has occurred, it is likely that the drivers that have crowded themselves onto the highway have forced such short headways that the speeds consistent with these headways drop off sharply. This is a condition that requires intercession on the part of the control system, so that the number of drivers permitted access to the freeway will not glut it to the point of a level of service below the maximum value.

Because a reduction of volume calls for an examination of speeds to see whether or not controls need be activated, another descriptor has been sought that may be developed from long-term data that will serve as a single unambiguous indicator. Such a descriptor is found in average lane density, expressed as the number of vehicles per lane per mile. It can be seen that volume rate is the product of speed and lane density, therefore density may be computed at intervals of time by dividing average volume by average speed. The curve of volume versus density also is roughly parabolic in shape and the critical density will be the maximum volume provided by critical speed (Fig. 2). In order for the control system to determine that the highway is serving as well as it is able, it is necessary only to monitor density—so long as it is at or below critical value the highway is operating efficiently; when critical value is exceeded steps should be taken to relieve the demand, such as restricting access along the route.

The foregoing discussion applies to the freeway under "normal" conditions.

However, when flow is at maximum volume the stream tends to be unstable. Perturbations can cause a reduction in service at one section, which then becomes a "bottleneck." The oncoming traffic quickly slows up and forms a moving queue, the tail of which progresses upstream as a "shock wave." This shock wave will propagate upstream so long as the volume of oncoming traffic exceeds the volume being passed by the bottleneck.

The rate of growth of the queue (speed of the shock wave) can be dependent on several factors. First, rate of flow through the bottleneck will determine whether or not the queue continues to grow or recedes. The oncoming volume, the road condition, and the geometric characteristics of the highway itself all affect the movement of the end of the queue.

The approach that has been taken to predict travel time when such a queue threatens to form is, first, to acquire information on the current behavior of traffic on the freeway. From the current behavior and a knowledge of the shape of experimentally derived volume-density curves, envelopes of maximum performance to be expected (volume vs density) can be constructed periodically in real time. These maximum performance envelopes are the only criterion to be used for traffic control decisions.

If it is determined that a bottleneck exists, the volume being passed by the bottleneck is compared to the oncoming volume upstream. In the event that formation of a queue is indicated, the oncoming volumes are integrated

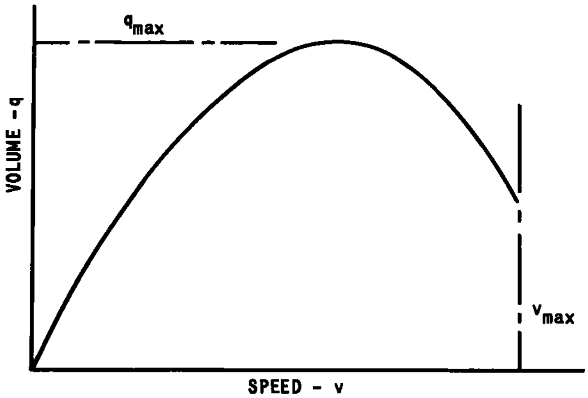


Figure 1. Typical speed-volume relationship.

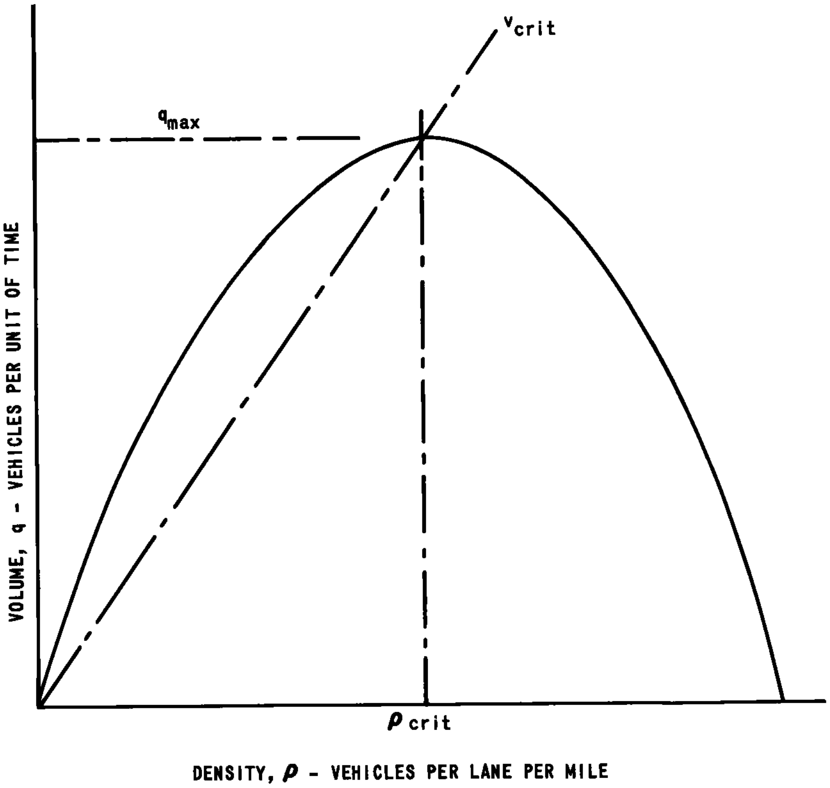


Figure 2. Typical speed-volume curve.

into a length of queue in which the vehicles are moving at the speed specified by maximum density that yields the bottleneck volume. Then, from the length of queue, the travel time is computed for vehicles, presently at some upstream location, to pass through the bottleneck.

IMPLEMENTATION OF TECHNICAL APPROACH

The site at which the experimental data were acquired for validation of the prediction logic is within the 3.2-mile length of the John C. Lodge Freeway that is under closed-circuit television surveillance. Although data to exercise the computational logic are being recorded automatically, correlation is based on actual happenings recorded by the television and human observers.

An assessment of the kinds of equipment available for the automated collection of data on traffic behavior led to the selection of over-the-road detectors for vehicle presence in individual lanes. Other means of vehicle detection would have yielded more accurate measurements of individual speeds, but one would have required an installation in the pavement and the other was a concept using some presently available equipment but not yet validated by tests. Also, the detectors selected were compatible with other instrumentation that NPG had purchased for the experimental segment of the freeway.

Accurate speed measurements were desired because, as previously noted, they were needed to construct curves of volume versus lane density. Instead, values of volume versus time occupancy have been calibrated to give average speeds, from which volume-density curves are constructed.

A detector was located over the center of each of three lanes at five overpasses along the freeway, for a total of fifteen. A layout of the region is shown in Figure 3. The detectors were capable of discriminating between "low" and "high" vehicles, the classification line being at a height of 6 ft. The data from the detectors were fed to a digital computer. Each sensing consisted of a pulse that lasted as long as the vehicle was under the detector. The pulse was tone coded as either "high" or "low."

In the computer, all vehicles lower than the classification height were assumed to be passenger vehicles and their times under the detector were added into one "bin" in the computer storage. All vehicles above the classification height were categorized as commercial and their times under the detector were stored in another bin. A subroutine in the computer recognized the start of each new pulse as the arrival of a vehicle and stored the counts in the appropriate classification for each detector.

Volume and occupancy counts were summed over an increment of time and the sums were recorded on punched paper tape. The precise manner in which the travel time predictions are made using the data is shown in Section (1) of Appendix A. The discussion is summarized in subsequent paragraphs of this chapter.

Observation and recording of information from the

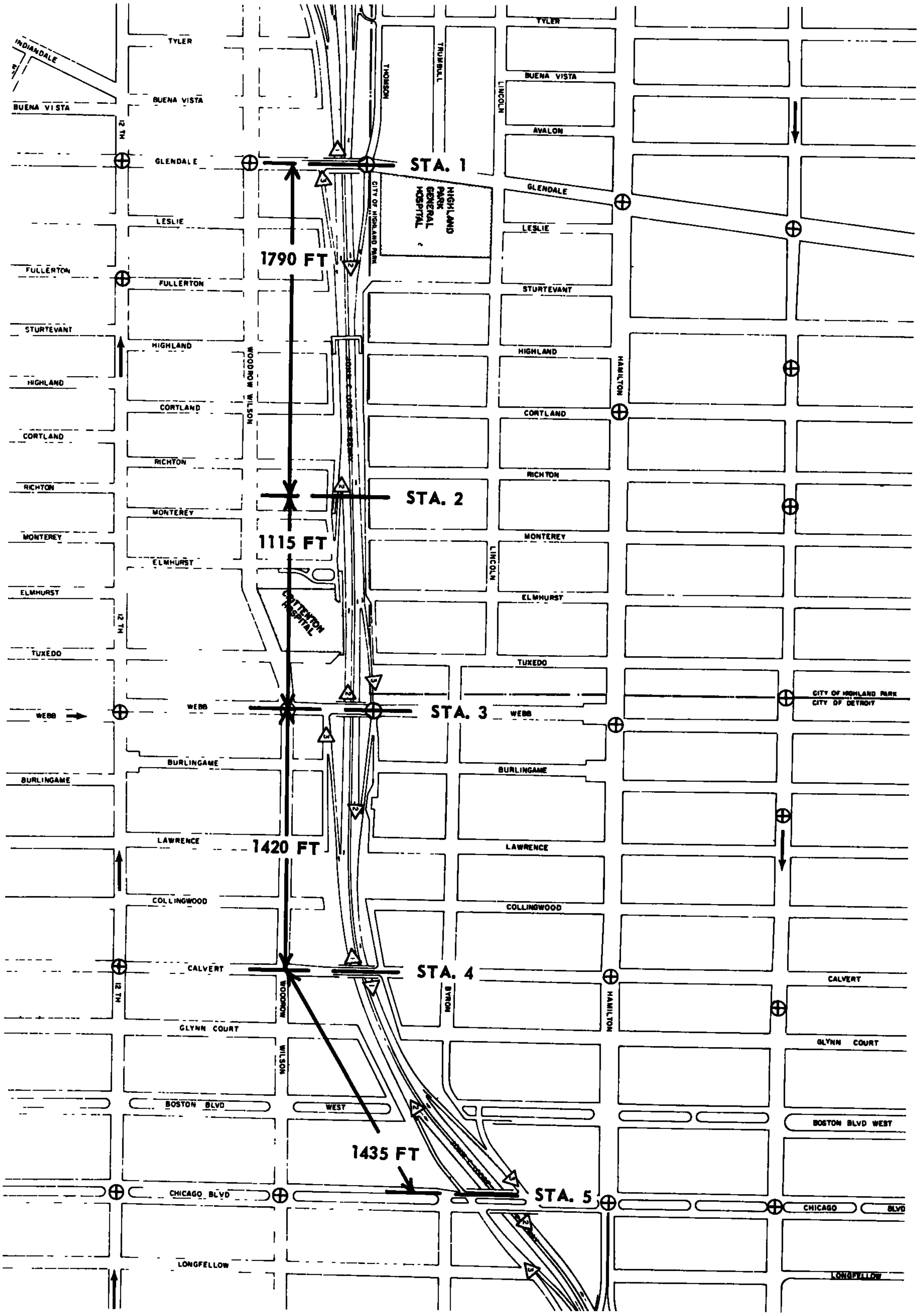


Figure 3. Instrumented portion of John C. Lodge Freeway, Detroit.

television monitors was done by NPG personnel. They were alerted to look for and record the time of occurrence of each traffic tie-up that caused an appreciable reduction in volume and speed at the downstream station of the five. For this cue, the time at which an abrupt reduction in volume was presented on an oscilloscope was taken as the time at which the bottleneck was established. Because the information that drove the oscilloscope presentation was furnished by the automatically collected and processed data, this display was used even though it might be apparent to the observers watching the monitors that the traffic tie-up had occurred sooner. Correlation with the automatic warning system required that travel time be computed from the moment when the prediction logic sounded the "alarm."

DATA PROCESSING

The geometric and other environmental characteristics of the freeway vary along its length so that there is reason to believe that the maximum performance envelope for traffic also will vary along its length. Therefore, the data taken at each station were processed to give the performance envelope for that station, the envelope being a volume-density curve representing the driver-established performance that may be expected at that station. Furthermore, because the motorists' driving habits and the environmental conditions change with time, provisions were made to keep the maximum performance envelopes on a current basis.

The data received were logged at 1-min intervals. High and low vehicle counts for the three lanes at each station were summed separately. Each count was multiplied by an average value for the length of commercial or passenger vehicles and the sum was divided by the total time occupancy to give the average speed, v_{ave}, for the group of vehicles; that is,

$$v_{ave} = \frac{\Sigma L \times \bar{L} + \Sigma H \times \bar{H}}{\Sigma O} \qquad (1)$$

in which

ΣL = the number of "low" vehicle counts for the time interval;

ΣH = the number of "high" vehicle counts for the time interval;

ΣO = total time occupancy during the time interval:

$\bar{L}$ = average length of a "low" vehicle; and

$\bar{H}$ = average length of a "high" vehicle.

Then the average lane density, ρ_{ave} was computed from

$$\rho_{ave} = \frac{\Sigma L + \Sigma H}{l\, v_{ave}} \qquad (2)$$

in which l is the number of lanes at the station.

Twenty points of volume versus density were computed for each station and used to construct an initial parabola. The parabola was forced through the origin and its coefficients were determined by a least squares fit. Figure 2 is characteristic of each such curve.

As a new point was computed for each successive minute, for each station, it was added to the initial group of 20 points. As each new point was computed and added to the group, the earliest one was dropped. Thus, the envelopes describing the expected maximum performance were updated each minute and represented the average traffic behavior for the past 20 min.

The decision to use 20 points, representing information acquired over a period of 20 min, was arbitrary. The actual variability of driver behavior with differences in driver population and changes in the environment is unknown at this time. Use of many points to construct the curve provides greatest accuracy and reproducibility of the function, but also makes the response "stiff" and sluggish if the situation is changing. On the other hand, use of just a few points allows flexibility and quick response to a changing situation, but also reduces the general applicability of the curve to the driver population because it is easily affected by atypical situations. An example of the need for quick response might be a sudden rainstorm or snow squall, which would change driving habits almost immediately. Such a response would not be desired if a "slug" of expert, tightly grouped, commuter drivers went through the freeway at the prevailing speed. Determining the best compromise for the number of points to construct the curves should be a matter of operational experience.

As long as the average speed at any station remains above critical speed, v_{crit}, the traffic is moving freely and nothing further need be done. However, if average speed falls below critical speed, average lane density is examined to see whether or not speed was reduced voluntarily by the last group of drivers or if it was forced down by high density. If ρ_{ave} is less than ρ_{crit}, no tie-up is indicated. But if the reverse is true (ρ_{ave} greater than ρ_{crit}), an indication of trouble downstream of the station exists.

(*Note:*—A comment is offered here concerning the use of both speed and density in the conditional inequalities in subsequent discussion. It has been stated that lane density is a sufficient criterion for expected performance maximums at any station. Remembering that the volume-density curves from which these performance values are read were constructed on a statistical basis, it is natural to expect that the distribution will contain points of higher performance. Thus, a unilateral decision that the freeway operation is deteriorating, due to a reading of density above ρ_{crit}, is avoided (no "false alarm") if it is found that the average speeds are continuing at or above v_{crit}.)

Yet another condition must hold before a warning of impending congestion is warranted. If the volume of traffic through the tie-up, as indicated by the current volume at the subject station, is sufficient to satisfy the demands upstream, it would be pointless to reroute the oncoming vehicles. They might suffer a temporary reduction in speed, or the condition might even clear before they arrived. Therefore, the third criterion that must be satisfied before the prediction logic is exercised is that the volume just upstream of the bottleneck section shall exceed that being passed by the bottleneck. In the ensuing discussion, Station 5 is treated as the downstream station in which the bottleneck occurs.

Therefore, the computer is programmed to make a prediction of travel time if, at any computing interval, the following conditions are found to exist:

At Station 5, $v_{ave} < v_{crit}$
At Station 5, $\rho_{ave} > \rho_{crit}$
Station 5 volume < volume at Station 4.

In the process of computing the travel time, several assumptions are made regarding the behavior of the drivers. None of these assumptions have been confirmed by any observation or experimental data. Indeed, it is doubtful if any applicable information exists. However, the assumptions seem reasonable in the light of ordinary human behavior. For instance, it is assumed that if a particular group of drivers are not maintaining maximum average speed over a segment of the freeway, they will not change their speed except when forced to reduce it. Another assumption is that vehicles arriving at the tail of a queue will reduce speed over a negligibly small period of time (infinite deceleration).

If a bottleneck is recognized as having occurred downstream of a station, the automated detection system will be unable to pinpoint the exact location. Therefore, an assumption must be made as to its location. Because the general likelihood is that the incident that causes the bottleneck may take place at random, the midpoint between detector stations is the best average location. Should some special geometric characteristic in a segment make another location a more likely trouble spot, the computer logic could be biased accordingly.

Another assumption is that no vehicles pass each other during the time that the queue is predicted to form. This point is touched on later in the description of the data processing.

Now, returning to the computational logic, as soon as it is found that the three conditional inequalities previously noted exist, the total number of vehicles in each segment between Station 5 and Station 1 is computed. This is an approximation; the length of each segment is multiplied by the number of lanes in the segment and the most recently computed average lane density. Variations in segment lengths and average speeds may require that an alternate approximation be used, a scheme that is discussed fully in Section (1) of Appendix A.

The data logging equipment at NPG utilizes punched paper tape and is speed limited to a read-out once every 60 sec, which may have contributed to some inaccuracy in determining the number of vehicles forecast to be in the queue. It is desirable that the segment lengths, d, be roughly equal to the product of normal average speed over a segment and the computing interval, or

$$d \approx v_{ave}\,\tau \tag{3}$$

in which τ is the computing interval. The real-time computer works at 20-sec intervals for read-out. Normal running speeds at maximum volume are about 40 mph (59 ft/sec). Thus, average travel for the 20-sec computing interval is about 1,180 ft. Average segment length (distance between instrumented stations) is about 1,200 ft, a compatible distance.

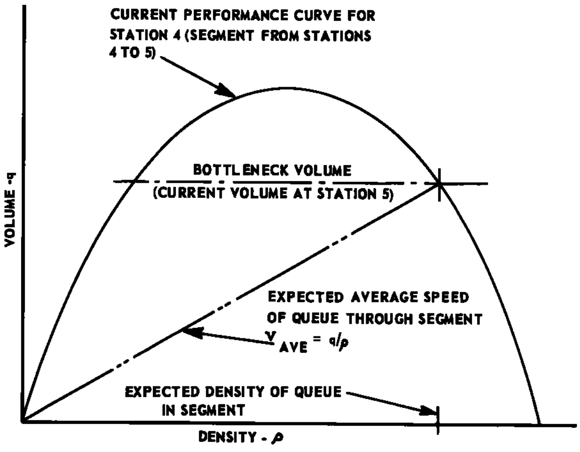


Figure 4. Traffic behavior mapped on volume-density curve.

The total number of vehicles between Stations 5 and 1 is "compressed" into a queue. First, the bottleneck volume is set as a horizontal line on the current volume-density curve for Station 4. It should cut the curve at two points. The higher of the densities at these two points is selected as the condition that should prevail in the segment between Stations 5 and 4 when the vehicles will have compressed into a queue and slowed down. Then, multiplying this density by the number of lanes and the segment length, the number of vehicles "stored" in the segment is determined. This procedure is illustrated in Figure 4.

The number of vehicles that can be stored in the segment is subtracted from the total number that was found between Stations 5 and 1. If there is a remainder, the same procedure is followed to compute the number of vehicles that can be stored in the segment between Stations 4 and 3. This number is subtracted from the remainder, etc.

When, finally, the position of the tail of the predicted queue is found, a correction is made to allow for the number of vehicles that will have moved through the bottleneck during this period of "compression." This is done by simulating an interception by a vehicle, presently at Station 1, with the moving tail of the queue. The vehicle is assumed to move at the current average speed computed for Station 1; the tail of the queue moves at the speed at bottleneck volume for the segment in which the tail is moving.

From the point of interception the vehicle is assumed to move at the reduced speed of the queue. Total travel time from Station 1 to the bottleneck location is computed. For this computation another assumption is made—that the bottleneck will persist for the period for which travel time is being predicted. This is another unsupported assumption that only operational experience can confirm or modify.

The prediction logic, as programmed for the IBM-7044, was a general approach to the problem. It assumed the existence of a large number of data collection stations and essentially one-dimensional movement of the traffic stream. Some special "stops" had to be written into the program to concede the fact that only five stations of data would be recorded. Also, several special quantities were included in the print-out to permit some hand computations to be made for travel time correlation. For instance, if a tie-up was indicated and all of the upstream stations were carrying greater volume than that at the bottleneck, an instruction ordered the computer to look for no more data, but simply to compute travel time from the upstream station to the bottleneck.

CORRELATION OF COMPUTED WITH OBSERVED TRAVEL TIMES

A page of travel time data taken from the television monitors by NPG personnel is shown as Figure 5.

The prediction logic using the data supplied on punched paper tape was programmed for the IBM-7044 machine. Exercising the logic for the five stations (15 detectors) required slightly more than 2 sec on this computer.

When the first tape of data was processed it was immediately apparent that conditions prevailing at Station 1 (Glendale) were too far removed from one-dimensional flow to allow the use of the data from that station. The source of the trouble is the access ramp from Glendale. Traffic entering the freeway from this ramp accounted for more than 10 percent of the volume at Station 2 (Monterey). Naturally, at near critical traffic volumes this entering stream would be expected to create a significant disturbance at the merge. The trouble was detected from the lower volumes, accompanied by lower speeds, at Glendale than at Monterey. No detectors were located on the ramps, so the count of vehicles merging into the stream (the approximate volume was determined by observation of the television monitor) was not entered into the data. The result has been that data from Station 1 (Figure 3) have been eliminated from these first runs of prediction correlation. Time remaining on the project was too short to allow for moving the Glendale detectors to another location or to visually count vehicles using the Glendale access ramp. It is felt that data from all five stations is a "must"; the loss of data from Glendale should cause a reduction in correlation accuracy, but to what extent is not known.

The punched paper tape data log, and the television-monitor-observed travel times covered periods during the morning commuter peaks. The first, on September 17, 1965, from 8:07 until 9:05, included 25 observations. The recording continued until October 8, for a total of 129 observations.

A typical print-out from the computer is shown as Figure 6, in which the downstream station of the four that were used qualifies as the bottleneck. The volume at all upstream stations exceeds the bottleneck volume so that all segments contribute to a growing queue. Travel time is computed from Station 2 to a theoretical bottleneck point 845 ft downstream of Station 5. The case is for the data given in Figure 5.

The print-out predicts that it will take 101 sec for a vehicle to traverse the distance, beginning at 7:54. An observation was begun at 7:54 at Station 1. The vehicle, a truck in lane 3, passed Station 2 at 7:54:40 and Station 5 at 7:56:41. Thus, the elapsed time from Station 2 to Station 5 was 121 sec. The time for the vehicle to travel the remaining 845 ft to the bottleneck was ratioed from its progress between Station 4 and Station 5, or $(845/1435) \times 42 = 24.7$ sec. Therefore, the "estimated" observed travel time from Station 2 to the bottleneck was $121 + 25 = 146$ sec. A similar observation 30 sec later for a passenger car in lane 1 gave the estimated observed travel time as 103 sec.

From the data collected, 46 such calculations of travel time have been made. Of this number, 14 used two observations each, accounting for 60 of the total of 129.

Thus, approximately 70 of the observations were not used. Many of these observations were lost because of the way the prediction logic is programmed. Once the computer recognizes a traffic tie-up and makes a prediction of travel time, its job is done; it has sounded the warning. So, it has been instructed to pass over any data for the region affected for the period of the predicted travel time. The observations made during the predicted travel times were not used because no data were processed to synchronize with them.

Computed and observed travel times have been correlated by determining the margin of error, in percent, of the time difference compared to the observed time. If two observations were made for a single computer print-out, the mean time was adopted as the observed time. Thus, using the foregoing example, the estimated observed travel time $= (146 + 103)/2 = 125$ sec, the time difference $= 101 - 125 = -24$ sec, and the error margin $= -24/101 = -24$ percent. The negative sign indicates that the computed time is less than the observed time. The results are shown in the form of a histogram in Figure 7, where each bar represents the number of cases that fell into that 10 percent range of error.

An additional bit of information resulted as a spin-off from the real-time solution of the volume-density curves. This was a comparison of critical speed, critical density, and maximum volumes for dry and rainy days (Fig. 8).

DISCUSSION OF INITIAL RESULTS

It is not felt that any statistically meaningful conclusions can be drawn regarding the results shown in Figure 7. The shape of the histogram cannot be classified into any of the commonly known distribution categories. Also, the center of gravity of the area appears to be located about at the null line, even though there are more errors on the negative side. Whether or not this asymmetry can be corrected by a bias in the mathematical program cannot be stated. Before any such conclusion could be drawn, it is felt that more data should be processed and in greater detail.

For one, the distance over which the observations and computations are made should be increased. It is reasonable (and instinctive) to expect that greater accuracy of

CORNELL AERONAUTICAL LABORATORY, INC.
PROJECT 3-2 - PREDICTING A TRAFFIC JAM

DATA SHEET

SHEET 1 OF 1

DATE: 9-28-65 DIRECTION: INBOUND

1. TIME OF LANE CLOSURE SIGNAL ACTUATION: NONE

2. TIME OF QUANTUM JUMP AT STATION 5: 0754

CUMULATIVE TRAVEL TIME (MIN : SEC)

STATION \ CLOCK TIME	0754		0754:30									
VEH. TYPE	3		1									
	LANE	T.T.	LANE	T.T.	LANE	T.T.	LANE	T.T.	LANE	T.T.	LANE	T.T.
1	3	0	1	0:0								
2	3	0:40	1	0:44								
3			1	1:06								
4	3	1:59	1	1:44								
5	3	2:41	1	2:11								

STATION \ CLOCK TIME												
VEH. TYPE												
	LANE	T.T.	LANE	T.T.	LANE	T.T.	LANE	T.T.	LANE	T.T.	LANE	T.T.
1												
2												
3												
4												
5												

CODE FOR VEHICLE TYPE:

VEHICLE	CODE
PASSENGER	1
PANEL OR PICKUP TRUCK	2
SINGLE UNIT OR VAN TRUCK	3
COMBINATION TRUCK	4
BUS	5

Figure 5. Sample observed travel time data sheet.

```
INDEX = 0
INDEX = 0
INDEX = 0
INDEX = 0
INDEX = 0
INDEX = 0
INDEX = 0
 THERE ARE  36. VFHICLES DOWNSTPEAM JAM STATION.
JAM CONDITIONS DOWNSTREAM  STATION    1 AT TIME   7 HOUR  54 MINUTES.   092865

VEHICULAR VOLUME FLOW IS NEVER LESS THAN THAT AT JAM STATION. INITIATING COMPUTATIONAL LOGIC.
 THE UNCOMPRESSED QUEUE LENGTH IS     4615.
THERE ARE  34.CARS IN SECTION    2  AT THE SAMPLE TIME   12
THERE ARE  40.CARS IN SECTION    3  AT THE SAMPLE TIME   12
THERE ARE  34.CARS IN SECTION    4  AT THE SAMPLE TIME   12
     0.760E 02   0.774E 02   0.726E 02   0.621E 02   0.        0.        0.        0.        0.
     0.          0.          0.          0.          0.        0.        0.        0.        0.
     0.793E 02   0.771E 02   0.751E 02   0.572E 02   0.        0.        0.        0.        0.
     0.          0.          0.          0.          0.        0.        0.        0.        0.
     0.206E-03   0.195E-03   0.195E-03   0.233E-03   0.        0.        0.        0.        0.
     0.          0.          0.          0.          0.        0.        0.        0.        0.
     0.277E-01   0.289E-01   0.257E-01   0.289E-01   0.        0.        0.        0.        0.
     0.          0.          0.          0.          0.        0.        0.        0.        0.
     0.206E-03   0.195E-03   0.195E-03   0.233E-03   0.        0.        0.        0.        0.
     0.          0.          0.          0.          0.        0.        0.        0.        0.
     0.252E-05   0.226E-05   0.230E-05   0.288E-05   0.        0.        0.        0.        0.
     0.          0.          0.          0.          0.        0.        0.        0.        0.
     0.144E-01   0.145E-01   0.133E-01   0.138E-01   0.        0.        0.        0.        0.
     0.          0.          0.          0.          0.        0.        0.        0.        0.
     0.          0.344E 02   0.402E 02   0.338E 02   0.        0.        0.        0.        0.
     0.          0.          0.          0.          0.        0.        0.        0.        0.
TOTAL CARS,CARS PER STATION,AT STATION  108.42      62.44 2
TOTAL CARS,CARS PER STATION,AT STATION  108.42      56.66 3
BPRIME,TC,TIC,CARNET HAVE VALUE, 3432.37     25.96      79.77      30.53
APPROX. LENGTH OF QUEUE IS      567.
MCRT, THERE IS NO STATION INSTRUMENTED UPSTREAM OF STATION K.
THE TIME REQUIRED FOR VEHICLES TO MOVE FROM THE LAST STATION THROUGH THE BOTTLENECK WILL BE 101. SECONDS.
RETURNING TO SAMPLING PHASE.

TIME =          2.01667
INDEX =15
INDEX =14
INDEX =13
INDEX =12
INDEX =11
INDEX =10
```

Figure 6. Sample predicted travel time print-out.

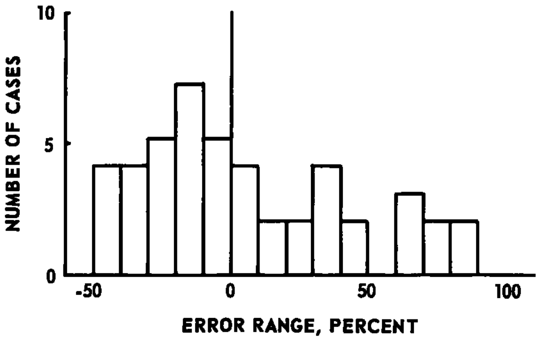


Figure 7. Presentation of results.

correlation should result if the travel times are taken over longer, rather than shorter, distances, at least from the standpoint of averaging out local departures from expected performance. The present data were derived for about 4,000 ft of the 3.2-mile length of the freeway that is under surveillance. Another 1,800 ft can be made available just by visually counting vehicles entering the freeway via the Glendale access ramp.

Another thing that may have contributed to the predominance of predicted times shorter than observed times was the selection of vehicles for visual timing. Of the 129 timed runs, 60 were for large trucks, buses, or tractor-trailer combinations, usually running in the curb lane. Normally, these vehicles represent the slow portions of the traffic stream. Although they have been used for nearly 50 percent of the data points, it is certain that they do

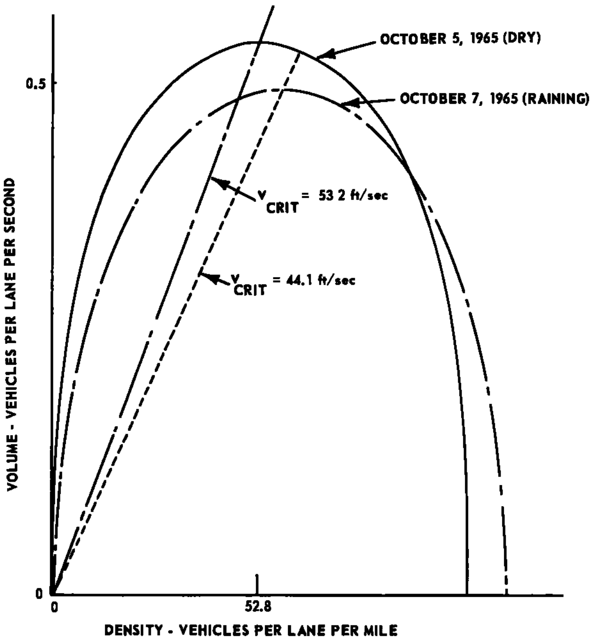


Figure 8. Volume-density curves for dry and rainy weather.

not represent nearly this fraction of the traffic stream. Although passenger cars are harder to identify, it is felt that such vehicles will furnish travel times that are more typical of the average movement.

A third point that remains very much in question is the selection of values for $\overline{H}$ and $\overline{L}$ to be used in the computation of average speed from time occupancy. This can be done only by making accurate measurements of the speeds of individual vehicles as they pass under the presence detectors. When a sufficiently large sample has been obtained of volumes, speeds, and time occupancy for each class, the average lengths can be computed and used thereafter as a standard. In the present case, values of $\overline{L} = 12.5$ ft and $\overline{H} = 40$ ft were used. Although the average length of passenger car is about 17.3 ft, use of this value in the computations would have led to absurdly high speeds for the known conditions on the freeway.

Although the mathematical model for the prediction logic is intended to be generally applicable, it is not necessarily true that its form is appropriate to the specific situation in which it is being tested. It was expected that some aerial observations would be made to see if the assumptions that (a) the bottleneck occurs at the midpoint of the segment in which it is located, (b) the bottleneck volume remains constant over the predicted travel time period, and (c) the location of the bottleneck does not change during the predicted travel time period, were borne out in actual freeway operation. If observations indicated a pattern of deviation from these preliminary assumptions, the mathematical model should be modified both for general application and for closer correlation in the present case.

Another problem is the data logging interval of 60 sec. As noted previously, it was expected that the computing interval would be 20 sec, to match the rate at which the computer performed its output calculations on the freeway. However, some shortcoming in the output equipment prevented the first group of tapes from being recorded at the 20-sec intervals. If this situation had been remedied, the data would have been more representative of current conditions.

The two parabolas of Figure 8 are not intended to be accurate representations of volume versus density over the entire speed range. Use of the parabolic form has been adopted for convenience of computation in defining traffic behavior in the critical range, near q_{max}. They do, however, bear out the findings of Malo, *et al.* (*3*) that environmental changes are accompanied by changes in the entire operating characteristic and not just by changes in the point of operation on a single curve.

During the manipulation of the processed data, in order to check out some apparent inconsistencies between concept and performance, it was found that the model included some unnecessary calculations. These had to do with the growth and dissipation of the queue. The "time to intercept" could be eliminated altogether, along with the calculations of "storage" capacities of the freeway segments. All that is needed to find travel time from any station to the bottleneck is to sum the number of vehicles between the station and the bottleneck and to divide it by the bottleneck flow rate. This time can then be compared with the time required for a vehicle at the station to traverse the same distance at the expected "unhindered" speed. The form of computation that was used resulted from a desire to reduce the number of iterative calculations to be made, but the latter may provide a simpler approach.

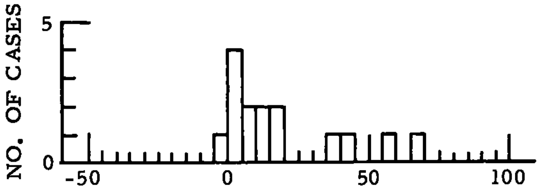


Figure 9. Presentation of additional results.

DISCUSSION OF FINAL RESULTS

The initial results presented for the 46 predictions of travel time as depicted in Figure 7 show significant errors when compared to the observed travel times. Upon review of the prediction logic used it was determined that the data rate used for computing lane occupancy was taken as 60 cycles per second, when in fact it was 40 cps. The large adjustment of passenger vehicle length, $\bar{L}$, from 17.3 ft to 12.5 ft would not have been required if the correct data rate of 40 cps had been used initially. In view of these developments, appropriate corrections were made and additional data were collected and processed, with the expectation that better correlation could be shown between predicted and observed travel times. In the final experiment, the error in occupancy data rate was corrected and a computer program was used for recording the surveillance data received from the computer at 20-sec intervals. The mathematical logic used in making the predictions is designed to be supplied with real-time surveillance data at 20-sec intervals.

Accordingly, NPG personnel recorded 30 observations of blockages during a period beginning February 10, 1966, and ending March 4, 1966. Punched paper tape records were made concurrently and an observer counted vehicles entering the freeway via the Glendale on-ramp. The latter count was needed in order to complete the volume records at the Glendale station.

The surveillance data were processed in the same manner as previously described. Unfortunately, between incorrect tape records and some malfunctioning detectors, 15 of the 30 observations could not be used. The final results of the remaining 15 observations are shown in Figure 9, presented as were the initial results in Figure 7.

It is immediately apparent, when comparing Figures 7 and 9, that the error range has been reduced drastically. The initial results produced errors extending from -50 percent to $+90$ percent. The current results show an error range from -5 percent to $+70$ percent. Although the small number of cases in the sample does not make for strong statistical significance, the trend is marked enough to show that the prediction program merits additional exploitation.

Specifically, the prediction logic should be programmed into the CDC-8090 computer at NPG so that real-time predictions can be correlated with travel times observed on the television monitors. Also, "false alarm" and missed prediction rates can be evaluated. Some "feel" for these rates has been gotten from the recorded data, but the method of computation does not warrant expressing them quantitatively. They appeared to be satisfactorily low, but truly acceptable evidence of the value of the prediction logic will best be gained by full-time real-time operation.

Another improvement that could be made in the computational procedure of the prediction logic would be to determine a more accurate equation for the envelope of volume-density performance. The parabolic shape that has been used might be replaced by a straight line radiating from the origin and transition into a parabolic or hyperbolic curve. Transition would occur at a density at which the drivers first would be affected by the proximity of other vehicles, at the posted or observed maximum speed.

CHAPTER THREE

THE AIRBORNE OBSERVER IN TRAFFIC CONTROL

Engineers are well aware of the progress that has been made in the design and capabilities of automated traffic control systems. Since the first fixed-cycle intersection signal replaced the traffic policeman, many devices to sense and control traffic movements have appeared. The increasing complexity of this equipment, however, indicates that the ultimate goal is to provide an automatic system which will exceed the ability of the human to assess and control traffic. Many observers feel that no available equipment has the flexibility, for instance, to optimize movement through an intersection as well as a properly trained policeman. The obvious reason for not employing more police as traffic controllers is the high cost involved and the fact that today's heavily traveled streets carry such high volumes during peak periods that a human controller would be "saturated." Also, the policeman at an intersection, even in a tower, has a fairly restricted view of the traffic complex. Therefore, he cannot evaluate the effects of his actions on adjacent areas of the complex.

Within the past decade there has been an increasing use of airborne observers for the purpose of reporting on traffic conditions. This is a very specialized application of the human traffic "controller." Usually the operation is conducted over a metropolitan area and in conjunction with commercial broadcast radio stations. From this vantage point the airborne observer can assess the general movement of traffic over a considerable region or can confine his attention to a seemingly troublesome intersection or two.

An initial study of this method of traffic control was reported in NCHRP Report No. 9 (*1*). This chapter describes the continuation and completion of that task to evaluate the effectiveness of an airborne observer.

The program to evaluate the effectiveness of an airborne-observer traffic control system developed as a result of auditing the broadcasts made by the radio station WEBR "Trafficopter 970." The content of these broadcasts seemed to make up two classes of control information, as follows:

1. For reduction of long-term (repetitive) traffic movement difficulties.
2. For reduction of spontaneous (accidental) traffic movement difficulties.

This study was concerned only with the airborne observer's effectiveness in the reduction of spontaneous traffic movement difficulties.

An airborne observer, trained to the task and familiar with the traffic complex he serves, can detect, evaluate, make decisions concerning disposition, and issue instructions on remedial or evasive actions to be taken when an incident causes, or threatens to cause, a severe traffic tie-up, particularly during peak periods. As yet, no automated system has this capability.

During the initial study, two relatively crude analyses were made of both the delay time suffered by motorists who were caught in traffic tie-ups and the delay time avoided by those who heeded broadcast advisories to take alternate routes. An estimate of costs was made to mount an airborne-observer patrol for 500 hr per year, using a helicopter service.

The results of the two analyses were that the motorists who used the alternate routes saved about $590 in one case and $640 in the other. This was based on $1.25 for the vehicle operation only; no estimate of the value of the motorists' lost time was attempted. The Trafficopter had guessed that such traffic emergencies, requiring the issuance of alternate route advisories, occurred at a rate of about 180 per year. An average helicopter service was estimated to cost $30,000 (one quote was as low as $22,500) and the observer would cost $5000. Thus, if the $35,000 cost was divided by 180 incidents, the savings had to be only $195 per incident to have the service pay for itself, assuming it to be tax supported. From the two incidents that were analyzed, it appeared that the service could justify its cost.

Data from which the two traffic situations were reconstructed for analysis were derived from the Buffalo Police and Fire Departments' and the Trafficopter's reports. A detailed consideration of the problem made it evident that any approach to accuracy in the analyses would require that the motorists themselves supply the bulk of the information on what happened, both among those who were delayed in the queues that were formed and among those who detoured via alternate routes that were suggested by the airborne observer.

PROJECT SCOUT

Personnel from this project, in conjunction with editorial and news reporters of radio station WEBR and the Buffalo *Courier-Express*, developed a publicity campaign to arouse public interest in Project SCOUT (Sky Control of Urban Traffic). The appeal to motorists was initiated by announcements over WEBR. These announcements were begun about the middle of April 1964. On Sunday, May 3, 1964, the *Courier-Express* carried a feature story and photographs of the Trafficopter in flight. The article repeated the need for driver assistance and described how the motorists could enroll in Project SCOUT.

The public was asked to request a traffic reporting kit, which contained a complete set of instructions on the use of three data reporting cards that were included. The project was identified as part of the National Cooperative Highway Research Program being administered by the Highway Research Board. Response to the appeal for driver participation came directly from individuals and also through several group representations such as industries, social clubs, the Automobile Club (AAA), and Erie County and local governments. Some 5,800 reporting kits were distributed in the Buffalo area during a period of about four weeks following publication of the newspaper article.

The data reporting cards were in the form of a questionnaire, postpaid and self-addressed for return. The information requested included times and locations for those drivers who were caught in a queue as a result of a traffic tie-up, the time at which they passed the bottleneck location, and similar information if they were held up on a cross street that was blocked by the queue on the street principally affected. Data requested from those who followed alternate routes included the time and point of departure from the normal route and the time and point of rejoining, as well as an estimate of the additional travel time caused by the detour. The instructions also noted that the cards were to be used only at the request of the airborne observer, because traffic incidents not controlled by him would not be included in the evaluation.

Going directly to the motorists with a public relations program to have them supply traffic data had been tried during the initial study of the project. At that time the response was not strong, but neither was the publicity campaign. The information acquired cost not much more than the expense of mailing data reporting cards and instructions to each driver, perhaps averaging $0.20 per driver. These are costs over and above the devising of the data acquisition program, which would have been done regardless of how the motorists were to be interrogated. To have duplicated the approximately 38,000 vehicle-miles that were reported by 332 drivers would have cost about $3,800 for the vehicle operation alone. The total travel time, about 1,500 hr, might easily have cost $2,250 in direct costs, if driver personnel could be hired at $1.50 per hour. The total, about $6,000, far exceeds the $60 or so that was spent.

No similar cost comparison can be made for Project SCOUT. It might have been possible to maintain an airborne patrol at all times when the Trafficopter was in operation. However, the patrol would have had to be prepared to take time-lapse photographs in color over the area affected by a traffic tie-up, the color to assist in identifying the movements of vehicles. From these photographs, vehicle delay time in the queues could be determined with good accuracy, but by a laborious method that would be expensive. The cost of the flight operation might be as low as $15 per hour, or $150 per week. However, this surveillance would not furnish any information on drivers detouring via alternate routes. It appeared to be impossible to discriminate vehicles following alternate routes, as the result of an advisory, from those ordinarily using the routes. Therefore, the dissemination of thousands of kits to "deputy researchers" was necessary in order to ensure that adequate numbers of these drivers would be present in the traffic to provide information needed for a reconstruction of the traffic situation when a tie-up occurred. The cost of making and distributing 5,800 kits was not more than about $2,000.

About 6 months after the beginning of the campaign, 120 cards had been received with data concerning 13 incidents. For several of these incidents there was enough information gleaned from the motorists, the police, and the Trafficopter to make possible an analysis of the delay experienced by the drivers involved in the tie-up and an estimate of the average delay avoided by those who took alternate routes on the advice of the Trafficopter. A description of the manner in which one of the analyses was made is included later in this chapter.

LIGHT AIRPLANE OPERATION

It was known that an airplane had been used for an airborne-observer operation (*1*) before any commercially available helicopter had been produced in the United States. (The first helicopter certificate was issued to Bell Aircraft Corporation in March 1946.) Also it was known that light airplanes could be flown for a small fraction of the cost of flying small helicopters, such as would be suitable for the traffic control mission. Therefore, it was of interest to this project to evaluate the usefulness of light airplanes for the airborne traffic control task, primarily with a view toward determining the minimum cost required to mount the airborne patrol.

A series of 25 flights was scheduled in which it was proposed that the patrols be flown at various altitudes ranging from 1,000 to 5,000 ft above the surface. (By Federal Aviation Agency regulation the minimum altitude permitted over congested areas is 1,000 ft.) The flight personnel consisted of a pilot-instructor and an observer, who carried photographic equipment and a stopwatch to aid in acquiring data that might be useful for analyzing traffic tie-ups for delay time. It was intended that the pilot would instruct the observer to fly the airplane at least well enough to determine if the task of reporting on traffic and controlling the airplane could be handled satisfactorily by one person.

The airplane patrol was used to continue the afternoon service for a period of 5 weeks, directly succeeding the normal discontinuation of the helicopter flights at the end of April. These 1-hr flights began on May 4, 1964, and ended on June 5, thus complementing the Project SCOUT campaign, which had been "kicked off" on May 3. Two traffic reporters from Radio Station WEBR alternated as observers. Each became proficient enough to maneuver the airplane as required to observe and report on traffic conditions. The airplane used was a Piper "Colt," a side-by-side seating, single-engine, 2-place, high-wing, cabin monoplane. It was flown in accordance with Federal Aviation Agency visual flight rules (VFR). No flights were missed due to bad weather during the 5-week period, although on several occasions the surface winds were up to 40 mph and on one occasion the ceiling was broken-to-scattered at 1,000 ft. The airplane was rented from a commercial operator for $10 per hour.

It was determined that at the minimum permissible flight altitude of 1,000 ft it was possible to see details well enough that broadcasts could be as complete as those made from the helicopter. Inability to hover or to drop to very low altitude eliminated the possibility of reporting on damage to vehicles in two cases where collisions occurred. (The value of these details for other than public appeal is questioned.) In clear weather and good visibility it was possible to report traffic conditions over at least a 3-mile radius when flying at 5,000 ft, especially when the view was unrestricted by buildings or trees bordering a street. In general it appeared that the speed of the airplane, which was varied from 70 to 90 mph, combined with the view from 2,000 ft, provided the right compromise between area coverage and observation capability of specific details. The final decision was to use 60 percent power to cruise at 85 mph at 2,000 to 3,000 ft, depending on ceiling and visibility. In the event of a call to the scene of an incident, cruise power was increased to 75 or 80 percent and the flight path was adjusted to reach an altitude of 1,000 ft on arrival over the scene. By this tactic, the airplane could respond more quickly to an emergency call than could the helicopter.

One difficulty experienced with the airplane was loss of VHF reception by the station when the airplane was orbited in a tight circle around a point. It was found that the radiation pattern from the antenna, which was mounted vertically below the airplane, was sufficiently flat so that the signal was momentarily lost when turns were banked in excess of 30 degrees. A proper combination of antennas might remedy this situation, but it was found that a 30-degree bank at reduced speed provided a tight enough circle so that a point could be observed continuously. Tape recordings of the rebroadcast transmissions revealed that background noise in the airplane was much lower than that in the helicopter (Bell Model 47). The side-by-side seating arrangement was not as convenient for vision out of the right side as a narrow cabin would have provided. Much of this difficulty could have been overcome by converting the cabin door to a full-length clear-vision panel. The view of traffic directly ahead, which is afforded by the "bubble" canopy of the helicopter, is lost to the light plane because of the engine installation. However, this shortcoming was easily overcome by placing the airplane into a "skid" or "slip" when such a view was desired.

Expected utilization of a light airplane for traffic control may be determined from published data (*4*), which show that VFR conditions prevail more than 80 percent of the time for most metropolitan areas in the United States. For the seven years reported (1938-1944), Buffalo weather was above VFR control zone minimums (ceilings of 1,000 ft or more and visibility of 3 miles or more) 82 percent of the time. Theoretically, flights may be conducted when the visibility is more than 1 mile, so long as the airplane is outside an FAA control area.

The use of a traffic researcher as aircrew resulted in the recording of low-speed lane density information by photography, the average speed being determined by elapsed time for a vehicle to traverse a known distance. These conditions always accompanied a tie-up, so that merging behavior and bottleneck volumes under emergency conditions could be observed. A camera and stop-watch should be carried at all times by an airborne traffic controller in order to record information on traffic, both for its value to traffic engineers and its interest to the local motorists.

COMBINED HELICOPTER-AIRPLANE OPERATION

An airborne patrol of 500 hr per year can be mounted at minimum cost by the use of both helicopter and airplane, with the latter serving as the primary vehicle. Thus, the observer would be airborne for an hour in the morning and another in the afternoon, five days a week, except for 10 holidays during the year.

Assuming that an airplane could be used for 80 percent of the flying time and that it would be rented for not more than $10 per hour, the annual cost would be $4,000. A helicopter plus pilot can be rented for $60 per hour, or less, from a number of flight services to whom inquiries were directed. Using the $60 figure, the annual cost for the remaining 100 hr would be $6,000. Thus, the total annual cost for 500 hr per year for aircraft service would be about $10,000.

It is expected that maximum benefits will accrue to the motorists from the airborne observer service if it is supported in such a manner that all radio stations in a metropolitan area will rebroadcast the traffic reports. This means that the service probably should not be sponsored by a commercial radio station (as a number of them are), but by local government or by commercial advertising. In the latter case, no direct cost for the service would be shown. If the operation is tax supported, however, it is estimated that the cost of the pilot-observer might be as high as $10 per hour, including fringe benefits and overhead. The pilot-observer and one or more alternates probably would be selected from among the personnel of a police or sheriff's department. In any case, flight pay for the service would not be expected to exceed $5,000 on an annual basis.

It is reasonable to suggest that the reports by the traffic observer would be rebroadcast by local radio stations at no direct cost to the public. The radio stations either would offer this as a public service or would find commercial sponsorship for the time. (A nine-station network in Detroit is planning to cooperate with traffic engineers to keep the public informed of traffic conditions in the city. A commercially-sponsored airborne observer service in Philadelphia is "aired" by six radio stations in that city.) Inasmuch as the Federal Communications Commission requires that all commercial radio and television stations support some programming in the public interest, the foregoing is realistic and not just wishful thinking. Similarly, including the airborne observer in a police radio network would involve no additional cost to the ground system. The expense not yet mentioned is that for two VHF transceivers for the aircraft, which would be about $2,000 for purchase and installation. One radio would operate at police frequency and the other at the mobile transmitter frequency assigned to commercial broadcast stations. The radios can be installed so as to be quickly interchangeable among substitute aircraft.

The system would be established on the basis that whenever possible all flights would be made with the airplane. The helicopter would be placed on a standby status only when weather predictions indicated that "below VFR minimums" of ceiling and visibility could be expected at the time a flight was scheduled. The fairly high reliability of the Weather Bureau in predicting these conditions would ensure that the helicopter would not be tied up needlessly for back-up purposes.

Thus, the total direct cost of mounting an airborne observer patrol might be as high as $15,000 annually, plus an initial investment of $2,000 for airborne radio equipment. (Maintenance costs on the latter would be negligible.) From the viewpoint of cost-effectiveness, this would be the expenditure that a practicing traffic engineer would have to compare with delay time saved in order to justify to the public the implementation of such a system on a tax-supported basis.

METHOD OF ANALYSIS OF AN INCIDENT

The general method for analyzing an incident was to try to reconstruct the traffic situation from the data received, and from the reconstruction to estimate the following:

1. The delay suffered by the drivers who were enmeshed in the traffic tie-up.
2. The delay that might have been suffered by drivers who would have been involved in the tie-up had not the airborne observer issued a warning and instructions on how to proceed to avoid the tie-up.
3. The additional travel time suffered by the drivers who

used alternate routes as a result of the airborne observer's warning and instructions.

Subtracting the sum of items 1 and 3 from item 2 gives the net saving in vehicle delay time realized by the motorists through the intercession of the airborne observer.

As had been noted previously, only the first item can be determined with reasonable accuracy by surveillance means requiring no participation by the drivers (time-lapse aerial photography). The second item must rely on knowledge of how traffic moved through the scene of the cause of the tie-up, as well as an assumption of traffic volumes on all routes affected by the tie-up. A further assumption must be made on how the traffic might have behaved in the absence of advisories from the airborne observer. The last item must be based wholly on written or verbal communications from the drivers who evaded the tie-up as a result of the advisories, the accuracy of the estimate being somewhat proportional to the percentage of responses.

The motorists who joined Project SCOUT were the primary source of information for items 1 and 3, although the Trafficopter assisted materially in locating the tail of the principal queue, at a given time, and also estimated movement through the bottleneck. As a rule he could offer only a spot check, because he continued on his patrol route after observing conditions at the site of the tie-up. Typical data cards returned by motorists are shown in Figure 10. These had to do with the Buffalo Skyway tie-up of July 21, 1964, for which the analysis is given in a subsequent section. Normal volume counts and average travel times for the route or routes affected by the tie-up were determined by project personnel or obtained from Buffalo's Division of Safety.

To estimate item 1, the growth and dissipation of the

IF YOU FOLLOWED A TRAFFICOPTER ALTERNATE ROUTE

1. Left regular route at ____________ intersection at ____ a.m. p.m.
2. Rejoined regular route at ____________ intersection at ____ a.m. p.m.
3. It took me ______ minutes to detour.
4. My total trip time was ______ minutes longer than usual.

IF YOU WERE CAUGHT IN THE TRAFFIC TIE-UP

1. I reached the end of the line of cars at 7:53 a.m. ~~p.m.~~
2. My location was NORTH END OF FR. BAKER'S BRIDGE (NORTH BOUND)
3. I passed the point that caused the tie-up at 8:17 a.m. ~~p.m.~~
4. I got across SKYWAY street at 8:20 a.m. ~~p.m.~~
5. I was stopped at:
   ______ Street at ______ a.m. p.m.
   ______ Street at ______ a.m. p.m.
   ______ Street at ______ a.m. p.m.

IF YOU FOLLOWED A TRAFFICOPTER ALTERNATE ROUTE

1. Left regular route at RT. 20 AND RT. 75 intersection at 8:22 a.m. ~~p.m.~~
2. Rejoined regular route at ELM ST. AND SWAN ST. intersection at 8:45 a.m. ~~p.m.~~
3. It took me 23 minutes to detour.
4. My total trip time was 10 minutes longer than usual.

JULY 21

IF YOU WERE CAUGHT IN THE TRAFFIC TIE-UP

1. I reached the end of the line of cars at ____ a.m. p.m.
2. My location was ____________
3. I passed the point that caused the tie-up at ____ a.m. p.m.
4. I got across ____________ street at ____ a.m. p.m.
5. I was stopped at:
   ______ Street at ______ a.m. p.m.
   ______ Street at ______ a.m. p.m.
   ______ Street at ______ a.m. p.m.

Figure 10. Typical data cards from Project SCOUT.

queue was charted from all available data. Rate of progress along the length of the queue and observed or estimated densities, or observed bottleneck flow, yielded the volumes as a function of time. From this information, the gross number of vehicle-hours of delay could be integrated. Subtracting the sum of the normal travel times from this total gave the net delay to vehicles caught in the queue.

Item 2 was estimated in the same manner, using total normal volume arriving in the queue. However, where the length of the queue extended to some "escape" intersection, it was assumed that the excess vehicles detoured of their own accord. Also, it was necessary, on occasion, to estimate a time at which the scene of an accident would be cleared and normal flow resumed so that the queue would dissipate.

The estimate of the number of drivers who used alternate routes was computed as the difference between the number of vehicles that actually passed through the queues subtracted from the normal volumes that used the route during the period of persistence of the tie-up. To this number of vehicles was applied the average value of the estimates of additional travel time required by the motorists who traversed the alternate routes.

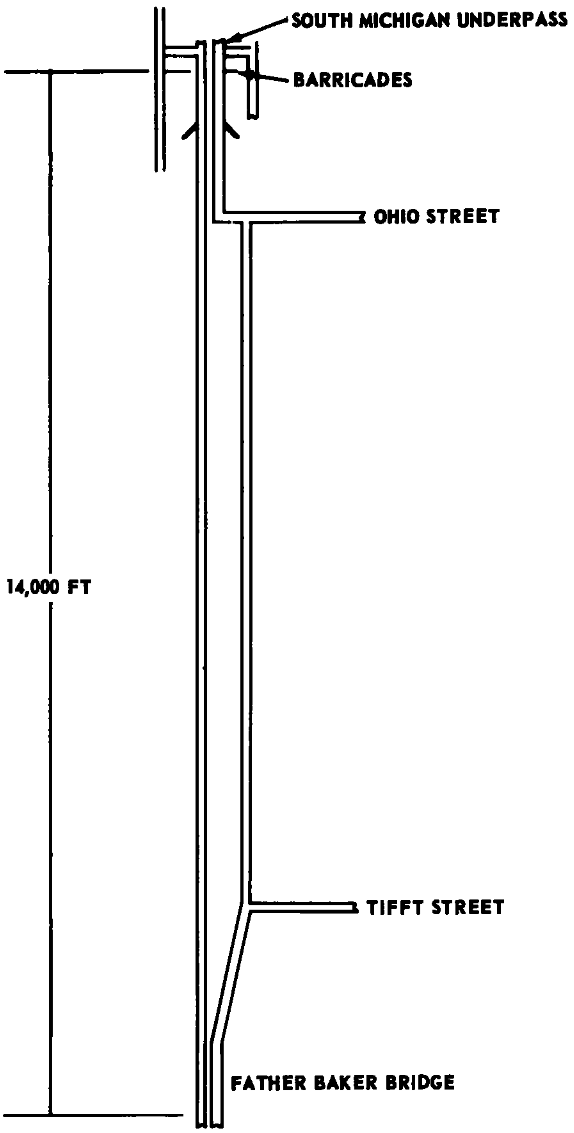


Figure 11. General plan of Buffalo Skyway route.

ANALYSIS OF A TRAFFIC TIE-UP

This analysis documents an effort to evaluate the effect of advisories to motorists on alternate routes, issued by the radio station WEBR Trafficopter, when construction barricades were placed across one lane of the Buffalo Skyway northbound on the morning of July 21, 1964. Although the airborne observer arrived on the scene about 55 min after the estimated start of the tie-up, he immediately announced the difficulty and advised mortorists to use alternate routes. The barricades were located about 500 ft south of the new South Michigan Street underpass (Fig. 11).

Log of Information Received

Trafficopter:

7:55 First report of trouble. Right-hand lane blocked about ¼ mile north of foot of bridge (see Fig. 11). Line extended to top of Father Baker Bridge. Traffic at a standstill from Ohio Street to bottleneck site.
8:01 Estimated delay 20-30 min.
8:25 Bottleneck flow reported as 20 passenger cars per minute. When a truck had to accelerate to pass through the site, only 8 vehicles would get through in that minute. Line contained 32 large tractor-trailer units.
8:26 Line shortened to Tifft Street.
8:38 Reported a 3-car smash-up, right in the middle of the jam. "Not causing a delay because traffic isn't going anywhere, anyway."
8:44 Accident site cleared.
8:48 Line still almost to Tifft Street. "Police should have re-routed traffic at Ridge Road."

Project SCOUT airplane:

9:37 Bottleneck flow counted as 100 vehicles in 5 minutes. Line still contains 120 vehicles.

Project SCOUT data cards:

1. At end of line at 7:58, passed barricade at 8:21, traveled 11,600 ft.
2. At end of line at 7:50, passed barricade at 8:15, traveled 11,600 ft.
3. At end of line at 7:53, passed barricade at 8:17, traveled 11,600 ft.
4. At end of line at 7:50, passed barricade at 8:15, traveled 11,600 ft.
5. At end of line at 8:25, passed barricade at 8:55, traveled 11,600 ft.
6. Alternate route took 20 min longer.
7. Alternate route took 13 min longer.
8. Alternate route took 5 min longer.

9. Alternate route took 10 min longer.
10. Alternate route took 2 min longer.
11. Alternate route took 5 min longer.
12. Alternate route took 15 min longer.
13. Alternate route took 0 min longer.
14. Alternate route took 15 min longer (tie-up on Thruway).
15. Alternate route took 10 min longer.

TRAVEL TIME DATA FROM PROJECT SCOUT

Five data cards carried consistent information on time to traverse the queue. In four cases the times of arrival at the tail of the queue were within an 8-min period, and travel times were 23, 25, 24, and 25 min. In the fifth case the travel time was 30 min. The average of these travel times, 25.4 min, over a distance of 11,600 ft, yields an average speed of $v_{ave} = (11{,}600 \times 60)/(5{,}280 \times 25.4) = 5.2$ mph.

The average delay suffered by those who used alternate routes is computed from the information on 10 data cards as: $75/10 = 7.5$ min.

RECONSTRUCTION OF THE TRAFFIC SITUATION

At 7:55 AM the line extended to the top of the Father Baker Bridge. From a map of Buffalo and the description of the incident, the length of the line is estimated as 14,000 ft (2.66 miles).

Flow rate at the bottleneck was counted as 20 vehicles per minute at 8:25, reducing to 8 vpm when a large tractor-trailer had to accelerate up the grade to pass the bottleneck. There were 32 such vehicles in the line at 8:25. The 20-vpm rate was confirmed an hour later.

The problem posed by this variation in bottleneck volume is that average lane density and average bottleneck flow must be estimated by the use of assumptions that cannot be confirmed. Certainly, the average speed of 5.2 mph appears to be a dependable figure, as does the bottleneck flow rate of 20 vpm for passenger cars. The 8-vpm flow rate was an isolated case, and was not confirmed by counting, although the observation was made that the presence of large trucks materially reduced the bottleneck rate.

Using the 20-vpm bottleneck rate at 5.2 mph yields an average density of $\rho = (20 \times 60)/(2 \times 5.2) = 115$ veh/lane/mile. For the 8-vpm flow rate the density would be $\rho = (8 \times 60)/(2 \times 5.2) = 46$ veh/lane/mile.

Intuitively, it seems that the density of 46 veh/lane/mile is much too low for a speed of 5.2 mph. An analysis of a similar incident that occurred in the southbound lanes of Fuhrmann Boulevard on May 18, 1964, showed that the photograpically determined density of 130 veh/lane/mile occurred at a speed of 6.8 mph. Normally, one would expect that the density at 5.2 mph would be even higher. The presence of the 32 trucks might account for a lower density, computed to be 115 veh/lane/mile at a flow rate of 20 veh/min.

A way to arrive at a reasonable estimate of density is to compute the rate of queue build-up as a function of bottleneck flow and correlate the time required to the probable time of placing of the barricade. If the density is 115 veh/lane/mile, the number of vehicles in the queue at 7:55 was $N_Q = 115 \times 2 \times 2.66 = 610$ veh. The number of vehicles that would occupy this distance under normal conditions is computed from volume counts and average speed data acquired by this project; that is, volume $= 1{,}850$ vph (7:30 to 8:30 AM), and average speed $= 35$ mph. However, this volume is for two lanes of traffic; therefore, $\rho_{ave} = 1{,}850/(2 \times 35) = 26$ veh/lane/mile and the number of vehicles in the 2.66-mile length of road would be $N_{normal} = 26 \times 2 \times 2.66 = 138$ veh. Based on these values, the additional vehicles required to form the queue would be $610 - 138 = 472$ veh.

If these vehicles were "recruited" by computing the difference between the normal flow (1,850 veh/hr) and the bottleneck rate of 20 veh/min, the build-up would appear to take 43.5 min, which would place the start of the queue at about 7:10 AM.

If the density is 46 veh/lane/mile and the flow rate is 8 veh/min, the number of vehicles in the queue is $N_Q = 46 \times 2 \times 2.66 = 244$, the additional vehicles required to make up the queue is $244 - 138 = 106$ veh, and the build-up would take $(106 \times 60)/(1{,}850 - 480) = 4.64$ min, which would have had the queue start at about 7:50 AM. This situation appears to have been impossible, because the queue was reported to be at constant length of 11,600 ft over a 5-min period before the Trafficopter advisory went over the air.

The contractor on the construction was called. He stated that he thought that the regular shift started at 8:00 AM, but that the barricades may have been set at 7:30. If this was the case, the queue can be estimated to have built up in 25 min. Then

$$\frac{(N_Q - 138)\ 60}{1{,}850 - q_b} = 25 \text{ min} \tag{4}$$

in which

$N_Q = 2 \times 2.66\ \rho$; and
$q_b = 2 \times 5.2 \times \rho$.

Solving Eq. 4 for density gives $\rho = 94$ veh/lane/mile. Based on this density, the average bottleneck volume during the build-up would be 16.3 veh/min, which seems reasonable in view of the observed performance. Also, this assumed starting time has a better "feel" to it than either of the other two computed starting times.

DISSIPATION OF THE QUEUE

At 8:26 the line had reduced in length so that the end was at Tifft Street; it was 2,400 ft shorter. In the 31 min from 7:55 to 8:26, the number of vehicles in the queue had been reduced by 85, the number contained in 2,400 ft. Therefore, the flow into the queue was computed to be $60\,(16.3 \times 31 - 85)/31 = 813$ vph, and the number of vehicles that were shunted onto the alternate routes by the Trafficopter's advisories, is computed to be $(1850 - 813)31/60 = 535$ veh for the period from 7:55 to 8:26.

It was observed that 120 vehicles still were in the queue at 9:37. Also, 100 vehicles had been timed through the bottleneck in a 5-min period, confirming the earlier count. A negligible number of vehicles was joining the queue.

Thus, it is estimated that the queue was entirely gone shortly after 9:43, a fact confirmed by the project engineer for the construction company on Fuhrmann Blvd. For this 77-min period, flow into the queue is computed as bottleneck flow less the decrease in number in the queue, or 60 [16.3 × 77 − (2 × 2.66 × 94 −85 −120)]/77 = 745 vph.

After 8:30 AM, the normal flow reduced to 1,230 vph, so the number of drivers who heeded the advisories during this period (from 8:26 to 9:37) is computed to be (1,230 − 745) 77/60 = 622, or possibly less. The vehicle count stopped at 9:00 AM. It would be reasonable to round off this figure to 600 vehicles. Then, the total number of drivers who heard the Trafficopter and followed his advice would have been about 535 + 600 = 1,135. The growth and dissipation of the queue is shown in Figure 12.

CALCULATION OF DELAY TIME IN QUEUE

Starting at 7:30 AM, it is assumed that vehicles passed through the bottleneck at an average rate of 16.3 per minute. Their average speed through the queue is taken to be 5.2 mph. Then the vehicles that joined the end of the queue at 7:55 will emerge (2.66 × 60)/5.2 = 31 min later, at 8:26 (coincidentally). Therefore, the number of vehicles affected by the 25-min build-up of the queue was 16.3 × 56 = 915, and each took an average of 15.5 min to traverse its length. Time spent in the queue was (915 × 15.5)/60 = 236 veh-hrs.

Similarly, a vehicle entering the queue at 8:26 emerged (11,600 × 60)/(5,280 × 5.2) = 25 min later, at 8:51; the number of vehicles affected was 16.3 × 25 = 408; and each took an average of 28 min to traverse the queue, for a time spent of (408 × 28)/60 = 190 veh-hr. Finally, the number of vehicles that spent an average of 12.5 min in the queue was 16.3 × 54 = 880, and their time was (880 × 12.5)/60 = 183 veh-hr.

Normal travel time for those vehicles that were in the queue is based on an average speed of 35 mph, as determined previously, and normal travel time = (236 + 190 + 183) 5.2/35 = 90.5 veh-hr.

Therefore, the net delay actually experienced by the vehicles that passed through the queue was (236 + 190 + 183) − 91 = 518 veh-hr.

POSSIBLE SITUATION—NO TRAFFICOPTER

It is difficult to predict what might have happened in actual fact had not the Trafficopter been available to issue advisories on the incident. From the rate at which the queue built up, it is reasonable to expect that no traffic was diverted until the advisory went out over WEBR. It must have been effective, because from that time on the length of the line receded, although the normally expected traffic would have increased the line, at least to some intersection where motorists might have turned off of their own accord. The approach taken, is to let the queue lengthen to the first main intersection upstream, then to hold it to that length as long as the approaching traffic exceeds the bottleneck rate. It is assumed that the overflow is penalized the same delay as noted by those who used alternate routes.

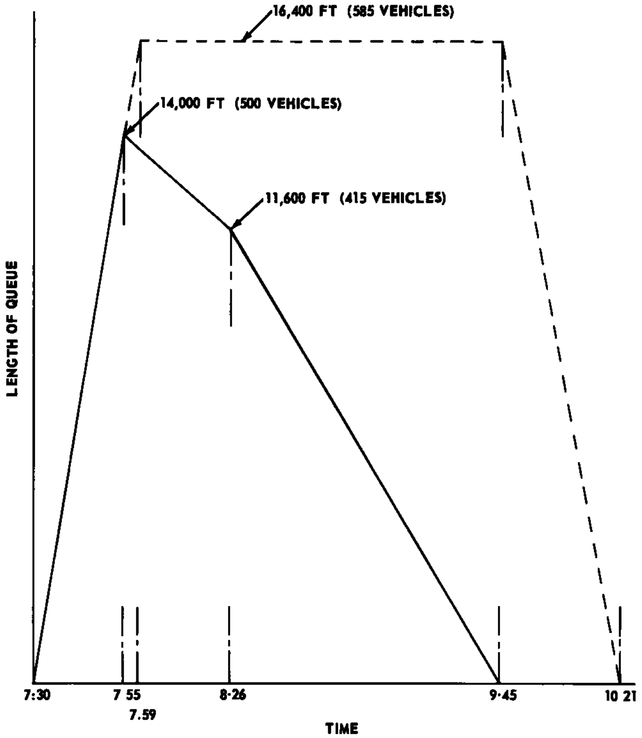


Figure 12. Comparison of time saving by use of Trafficopter with normal traffic tie-up.

The queue profile is based on the assumption that the bottleneck flow of 16.3 vpm is maintained for the duration of time required for the normal volume of vehicles to pass. The total number of vehicles affected is based on a flow of 1,850 vph up to 8:30 and 1,230 vph until 9:45, after which the flow is considered to be zero. The intersection that provided the last escape from Fuhrmann Boulevard was, at the date of this incident, at Ridge Road, about 4,800 ft south of Tifft Street. Therefore, it is assumed that the queue length would have remained static at 16,400 ft until 9:45, after which it would have receded at the bottleneck flow rate.

From the computed rate of growth of the actual queue it is estimated that it would have reached Ridge Road after 25 × 16,400/14,000 = 29 min after the barricades were placed at 7:30 AM, which would be at 7:59. The time for a vehicle to traverse this queue would have been (60 × 16,400)/(5,280 × 5.2) = 35.8 min. Then the vehicles that entered the queue just as it reached its full length at 7:59 would emerge at 8:35. During the time from 7:30 to 8:35, 16.3 × 65 = 1,060 veh would have spent an average of 18 min in the queue, for a travel time of 1,060 × 0.3 = 318 veh-hr. In the time from 8:35 until a vehicle that entered the queue at 9:45 emerged, which would be at 10:21, the travel time for 16.3 × 106 = 1,730 veh would be 1,730 × 0.6 = 1,038 veh-hr.

Normal travel time for the hypothesized queue is computed as before. Traverse time for the 16,400-ft distance is 8.9 min. Thus, the total travel time is $(1{,}060 \times 4.45 + 1{,}730 \times 8.9)/60 = 335$ veh-hr, and the net travel time lost in this hypothetical queue would have been $318 + 1{,}038 - 335 = 1{,}021$ veh-hr.

EVALUATION OF TRAFFICOPTER EFFECTIVENESS

The evaluation of the effectiveness of the Trafficopter advisories is based on the difference between what is estimated to have happened and what might have happened. It was computed that 518 veh-hr were lost in the actual queue, and that $535 + 600 = 1{,}135$ veh took alternate routes at an average delay of 7.5 min each. Therefore, the latter suffered a delay of 142 veh-hr.

The loss in the case of no Trafficopter was computed previously at 1,021 veh-hr. Thus, the benefit to the motorists was $1{,}021 - 518 - 142 = 361$ veh-hr, which, at a rate of $1.25 per hour for the vehicle cost only, is a saving of $450.

Figure 12 indicates the reconstructed queue length and time of persistence and the queue that it was estimated would have formed if the airborne observer had not exercised control.

RESULTS OF COST-EFFECTIVENESS STUDY

Altogether, nine traffic incidents have been reconstructed. In each an analysis has been attempted to determine the vehicle delay time avoided by motorists who heeded the Trafficopter's alternate route advisories. In three cases insufficient data were received to compute an over-all saving, although it was known from Project SCOUT questionnaires returned that a number of drivers had avoided the tie-ups, all of which were quite large, each involving about 1,000 vehicles.

In the other six cases the savings varied from $450 to $1,280, the average for all nine being $465 per incident. A log was kept of practically all traffic broadcasts over a 12-month period. During this period there were recorded 104 incidents involving traffic tie-ups, of varying degrees of severity, for which alternate routes were advised and other control information offered. If it is assumed that the average saving of $465 can be applied to all of these incidents, the annual saving would be $48,400, less the cost of the operation, which was estimated at $15,000 annually plus an initial cost of $2,000. It is emphasized again that the saving is based on vehicle operating cost only; no attempt is made to place a value on the time of the motorists who would otherwise suffer delay.

Therefore, it is concluded that an airborne observer service, when implemented as shown in this section and operating in a metropolitan area, can provide benefits to the public much in excess of its cost to that public.

It is not unreasonable to expect that the Buffalo operation could be much more effective than it is at present, inasmuch as radio station WEBR is the sole broadcaster of the Trafficopter's reports and is only one of four major stations in the area. No figures as to the relative sizes of the listening audiences were available to this project.

CHAPTER FOUR

PRELIMINARY INVESTIGATION OF AN AIRBORNE TELEVISION SYSTEM FOR TRAFFIC SURVEILLANCE

The concept described in this chapter springs from several sources, each contributing in either a philosophical or a technical manner. First, the value of the "overview" of traffic has been made evident by the airborne observer operation described in Chapter Three. The point is clearly established that area surveillance of traffic operations can provide information for a more powerful control system than one based on data collected at discrete locations. Second, it is inherent that any operation must incur losses, because no system is perfect. Therefore, it is a constant aim to minimize the numbers of personnel involved in any hazardous portion of an operation, in keeping with the cost-effectiveness required of the system. That the manned-aircraft airborne observer service includes an operational hazard is attested to by the fact that fatal accidents have occurred in Chicago and Houston in the course of the traffic helicopter flights.

Third, this project is principally concerned with the development of automated means for the surveillance, decision-making, and control of traffic on urban freeways and street networks. Experience with the closed-circuit video equipment used on the John C. Lodge Freeway in Detroit indicates the feasibility and practicability of using this type of equipment to provide area surveillance of traffic through appropriately located cameras transmitting the desired information to a ground-based control room.

By integrating these three factors, the conclusion was drawn that a useful long-term research tool for traffic would be a closed-circuit video camera maintained aloft by an unmanned tethered balloon. The altitude would be sufficient to give adequate vision range without loss of resolution due to masking at low deflection angles but not so great as to lose resolution due to small image size.

Figure 13. Interchange at junction of Lodge and Ford Freeways, Detroit.

A traffic surveillance system is suggested in which a closed-circuit television camera is borne aloft by a tethered balloon. This concept provides for long-term area surveillance of vehicle movements at a cost lower than that for a manned heavier-than-air system and without the attendant risk to personnel.

DESCRIPTION OF THE CONCEPT

Use of the closed-circuit video derives from experience with this type of equipment at the National Proving Ground for Freeway Surveillance, Control and Electronic Traffic Aids (NPG), located on the John C. Lodge Freeway in Detroit. Although the information presented by the video currently is used mostly in a qualitative manner, it is possible to make some measurements from observations of vehicle movements. This capability would be enhanced if the camera could be raised overhead and its information recorded on video tape. By raising the camera well away from the surface, masking and foreshortening are reduced for oblique photographs and it is possible to get planform presentations of significant areas. In the latter, the scale would be constant over the whole presentation. Video tape equipment now coming into the market for home use would provide capability for making permanent records at reasonable cost. Either general data or singular events could be studied at leisure, because the tape recorder could make successive tapes that either could be set aside for review or degaussed and used again in the recorder.

To reduce or eliminate involvement of personnel from the airborne operation, by using a 3-wire tethering arrangement, it is possible to get both position and orientation stabilization of the balloon in reasonable wind conditions. An additional feature, remote control of the camera field of view by a "pan" and "tilt" mechanism, makes it possible to operate the airborne system unmanned.

A preliminary investigation was made for such an installation at the intersection of the John C. Lodge Freeway and the Edsel Ford Freeway in Detroit (see Fig. 13).

SOME DESIGN CONSIDERATIONS

One of the important technical points to be settled is the altitude at which the balloon should operate. Other considerations aside, two opposing tendencies set the upper altitude limit. From the standpoint of area coverage it would be desirable to have the camera stationed high enough to allow viewing of the entire field of interest in a single planform presentation. This is not possible. First, camera resolution sets one limitation on altitude.

In use at NPG are the General Electric TE-9, a transistorized, positive interlace camera having a vertical resolution of 525 lines. Field of view with a 1.0-in. wide-angle lens is 28° × 41°. If the larger angle is used to observe a ¼-mile length of the interchange, the camera altitude would have to be 1,760 ft, assuming vertical viewing. The cameras are being used at ranges of about 1,200 ft along the John C. Lodge Freeway. At this range, resolution of the 5-ft height of the average passenger car is quite adequate. Therefore, assuming adequate color or shadow contrast with the pavement as a background, the 17-ft length of the vehicle should be readily resolved, the vertical and horizontal sharpness of the images being about the same. But the 6-ft width of the vehicle would be critical for viewing, so it is suggested that the 1,700-ft altitude previously mentioned would be a reasonable upper limit for the wide-angle field of view. It would be pointless to suggest a higher altitude for a telephoto lens, because the only advantage to be gained would be reduction of masking. The latter poses no problem at the viewing angle of 20.5° from the vertical.

Second, a moored balloon in the United States airspace is subject to regulation by the Federal Aviation Agency. Part 101 of the Federal Aviation Regulations, Subchapter F, contains the rules governing use of balloons. Normally, this would limit the top of the balloon to a height of 500 ft above the surface. However, being within a 5-mile radius of Detroit City Airport, the interchange site comes under the jurisdiction of a more restrictive altitude regulation. After examination of the geometry of the airport's instrument landing system clearance envelope surfaces and two existing buildings in the immediate vicinity, the Federal Aviation Agency determined that the altitude limit would be 400 ft above the surface.

In view of these considerations affecting operational altitude, particularly the 500-ft limit specified by the FAA, no other factors were investigated. However, for an installation of this type at some other location, the problem of video transmission without loss of resolution might affect the altitude limitation.

A third consideration is the wind velocity. A moored balloon is subject to drag forces due to the local wind and gust velocities. Therefore, the tethering wire geometry, wire size, balloon envelope, anchorages, etc., must be designed to provide adequate strength and stability. Because this particular installation would be for research purposes only, as contrasted with a full-time operational system, an upper limit of 20 mph wind velocity was chosen for design. This value would allow the balloon to be used for a large majority of the days in all seasons in the Detroit area.

Snow loads are not included in the design, nor is any other form of precipitation. It is expected that the non-wetting material of the envelope will shed water well enough so that the system can operate in light or moderate rain. Although it may be desirable to observe the behavior of traffic during a snowfall, it was felt that the additional lift requirement would be too harsh for the value of the data. Of course, the balloon could be cleared and lofted immediately after a snowfall so that the observations could be made on a snowy surface. Daylight operations only are assumed.

A final design consideration, the lifting gas to be used, was decided purely on a policy basis. Because the balloon would be moored over a heavily traveled road located in a densely populated area, the risk of fire was not to be countenanced. Therefore, although it involved a significant cost penalty, helium was chosen as the lifting gas rather than hydrogen or illuminating gas.

PRELIMINARY DESIGN OF THE SYSTEM

The design of the system is based on the use of a balloon having moderate life expectancy and reasonably low gas leakage. The latter is of some significance, because helium is a relatively expensive gas. The requirement for orientation and for stability in a 20-mph wind led to a three-wire tether geometry in which the wires are at a 45° inclination to the horizontal. A sketch of the installation is shown in Figure 14.

Balloon envelope materials examined varied in weight from a 0.0015-in. thick mylar at 1.5 oz/sq yd to a hypalon-dacron-capran laminate at 10 oz/sq yd. A mid-range material, neoprene-coated nylon at 5.5 oz/sq yd, was selected for fabrication of the balloon. The cable junction, shroud attachment and camera platform would be integrated into a single assembly, shown in Figure 15. The conical aluminum housing would shield the camera from the elements, as well as carry the tension loads due to the shroud lines and cables. Shroud lines would be nylon cord approximately ⅛-in. in diameter. The cables would be standard ⅛-in. diameter 7 x 19 stainless steel aircraft cable.

For an installation that would be stable regardless of wind direction, a 40-ft diameter balloon is required. If a prevailing wind direction is resolved, a wire can be installed in that direction and a 35-ft diameter balloon will suffice.

The cable that carries the electrical power to the camera also would contain the coaxial lead bringing the video back to the ground. Additional leads would be available to power the motors that would tilt the camera along two axes. This cable also would be used to raise and lower the balloon. Because the free lift of the balloon is about 1,100 lb, a powered winch or capstan would be required. Small, 24-v d-c motors would be used to tilt the camera, reversal being achieved by changing polarity of the power leads.

PRELIMINARY ESTIMATE OF WEIGHTS

G. E. TE-9 camera, 1- and 6-in. lenses, junction box	15 lb
Cables, mooring, ⅛-in. dia. stainless steel 7 × 19 aircraft, 1,485 ft	45 lb
Camera cable, 350 ft	131 lb
Suspension platform and upper and lower attachments	15 lb
Camera tilting mechanism and drive motors	6 lb

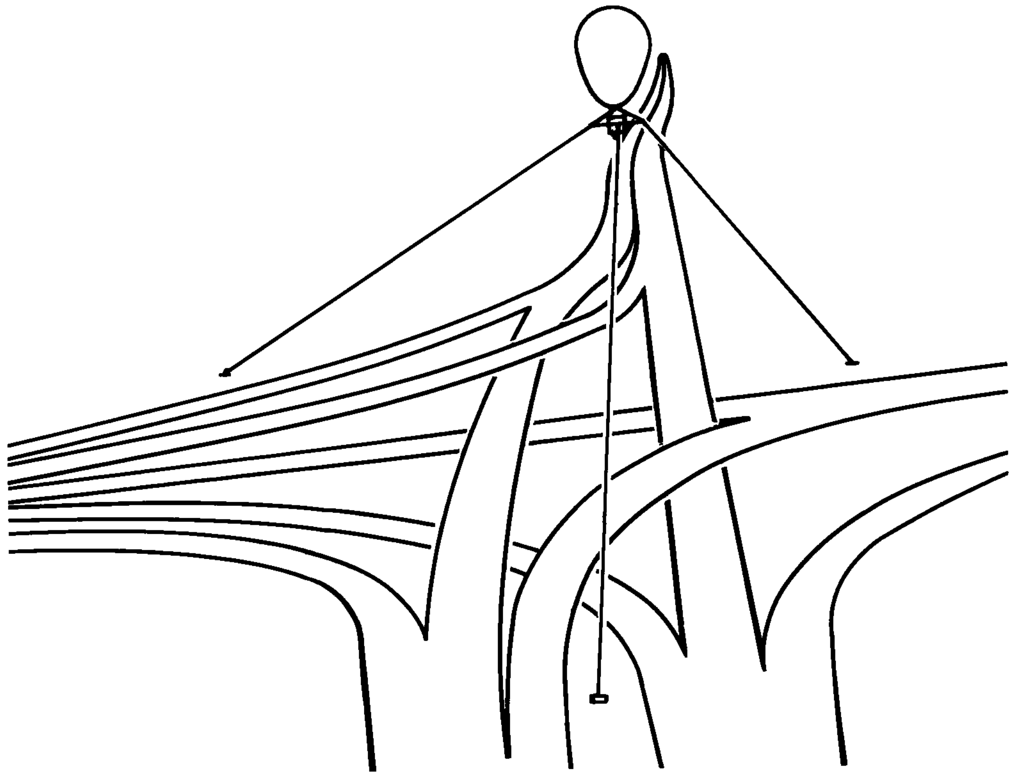


Figure 14. General view of proposed balloon installation.

Running light	5 lb
Balloon, 40-ft diameter	216 lb
Shroud	10 lb
Total	443 lb

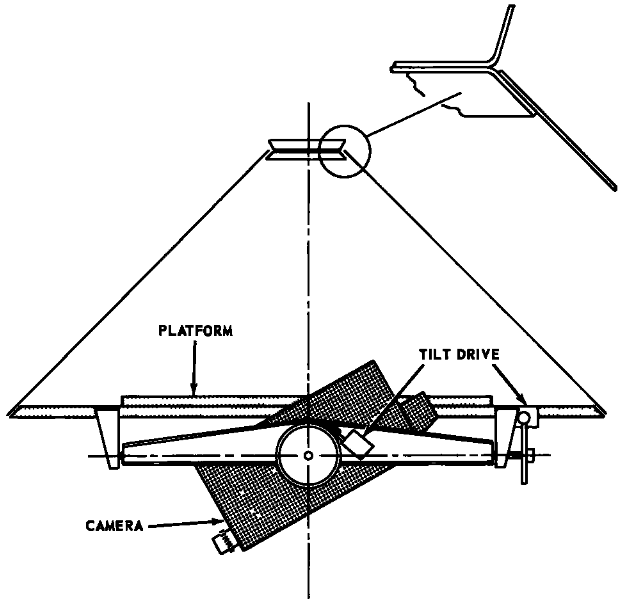


Figure 15. Camera support platform and tilt drive assembly.

The volume of a 40-ft diameter balloon, when in the natural balloon shape, is approximately 26,350 cu ft. When inflated with helium, the gross lifting force is 1,600 lb.

PRELIMINARY ESTIMATE OF COSTS

An estimate of the cost of the airborne system is made except for the price of the television camera and its cable. These items are not included because different types of cameras and lenses can have significant cost differences.

Aircraft cable, 1,500 ft @ 12.30 per 100 ft	$ 185
Suspension platform, camera tilting mechanism, and drive motors	2,000
Balloon, shroud lines, and running lights	9,000
Helium, 30,000 cu ft @ $7.50 per 100	2,250
Total	$13,435

Installation costs should be relatively reasonable. The only appreciable weight item to be transported is the helium, which is packaged in high-pressure bottles. A winch for raising and lowering the balloon would be rented from any towing service, or might be part of the city's ordinary garage equipment. A rough estimate of a cost to install augur-type tie-down rings in the turf to act as anchor points, and to perform other tasks necessary to get the balloon lofted, might be $1,000. After the initial installation, subsequent handling should be a nominal expense.

REFERENCES

1. WEINBERG, M. I., "Traffic Surveillance and Means of Communicating with Drivers—Interim Report." *NCHRP Report No. 9* (1964).
2. WEINBERG, M. I., ET AL., "Digital-Computer-Controlled Traffic Signal System for a Small City." *NCHRP Report No. 29* (1966).
3. MALO, A. F., ET AL., "Traffic Behavior on an Urban Expressway." *HRB Bull. 235*, pp. 19-37 (1960).
4. "Classified Flying Weather for the United States." U. S. Dept. of Commerce (Dec. 1946).
5. "An Introduction to Traffic Flow Theory." *HRB Spec. Rep. 79* (1964).

APPENDIX A

SURVEILLANCE LOGIC

1. MATHEMATICAL LOGIC FOR COMPUTING TRAVEL TIME ON AN URBAN FREEWAY

During the processing of data for Chapter Two it was noted that a major change in computational procedure should be made. A routine had been developed to estimate maximum queue length and then to follow the dissipation of the queue as vehicles from successive stations upstream approached the bottleneck point. This routine eliminated some iterative computations, but is not necessary. In fact, it appears that, even for a generalized application, the iteration is preferable to the queue length calculations. Therefore, the latter have been eliminated.

This section discusses a theory and presents a computational procedure for a logic to locate a bottleneck in a one-dimensional, originally steady-state, traffic stream. Whether or not this routine will have practical value will be ascertained from the characteristics of the traffic data used to implement the logic. The subroutine may be of academic interest only.

This description details the manner in which data collected by vehicle detection equipment on an urban freeway may be processed to make predictions of travel time on the freeway. More specifically, the prediction is made only in the event that an incident or accident of some sort causes a disruption in the normal movement of the freeway traffic. It is expected that a queue of vehicles is formed for which movement is hindered.

Recognizing that such a freeway when operating in a free-running condition can offer a much higher level of service to motorists, both individually and collectively, than can the surface arterial streets, it is desirable to maintain it in this condition as long as possible. However, unforeseen incidents are bound to occur that will cause a deterioration of the service on the freeway. No control system can prevent such a situation from developing, short of restricting the use of the freeway to a negligible service level.

The objective is to determine, in the event of an incident, the reduction of service level that will occur on the freeway, by a forecast of travel time, and to signify when the forecast indicates that the motorists would be better served by using the surface street system. The general task may be stated as follows:

1. To define that travel time along the freeway that can be equalled by travel along alternate routes (surface arterials).
2. To develop a traffic surveillance procedure and computational logic that will monitor and predict traffic behavior in order to implement controls to divert traffic from the freeway when the travel time threatens to exceed that in item 1.

This research is concerned with the second item. The clues to its behavior must be read from the traffic by detection equipment, and it must be within the capabilities of the equipment to do so. Furthermore, the data thus collected must be suitable for processing in a high-speed computer so that the control system may operate in real time.

HIGHWAY CAPACITY CRITERION

The time-honored criterion for traffic movement on highways has been volume flow, which is expressed as the number of vehicles passing a point per unit of time. Some postulate that the capacity of a highway be expressed in terms of volume, as defined; the higher the volume attainable, the better the road. Others maintain that volume alone as a measure is insufficient and misleading and that a highway operating at maximum attainable volume probably is operating at less than peak efficiency as far as service rendered to the motorists is concerned. An analytical proof of the correct hypothesis follows.

Two cases are considered. The measure of desirability is either:

1. To minimize over-all traverse time for a fixed number of motorists seeking to use a given highway; or
2. To maximize the amount of travel accomplished by the number of vehicles contained in a length of highway.

Travel Over A Specific Traffic Facility

In order to show that volume is a misleading criterion, a slightly idealized * representation of a very usual traffic situation is considered. Assuming a highway of length, L, over which a number of vehicles, V, are awaiting passage, if the vehicles are able to enter at a steady rate, q vehicles per unit of time, and can proceed along the highway at speed s, the over-all traverse time, T, for all vehicles to complete the passage is

$$T=\frac{V}{q}+\frac{L}{s} \qquad \text{(A-1)}$$

Under the assumption that the value of T is to be minimized, the best obvious solution to Eq. A-1 is for both q and s to be maximum values. But these two parameters are not independent of each other; as is well known, q is a function of s. Furthermore, the many curves of volume versus speed that have been drawn by traffic researchers from observed data show that q_{max} does not occur at s_{max}. In general, the appearance of the curves is as shown in Figure A-1.

By inspection, Eq. A-1 shows that if the number of vehicles, V, seeking entrance to the highway is large relative to its length, L, a point on the speed-volume curve close to q_{max} would be desired. This is the situation that obtains in a bottleneck such as a toll barrier, vehicular tunnel, or merge. If the converse is true (small V and large L), minimum T is obtained at a q and s near s_{max}. In any case, the solution is found by writing Eq. A-1 as

$$T=\frac{V}{f(s)}+\frac{L}{s} \qquad \text{(A-2)}$$

and differentiating, setting $dT/ds=0$, and solving for s in terms of V and L. It appears that the solution giving minimum over-all traverse time will *not* be such that $q=q_{max}$.

Steady-State Travel over a Continuous Highway

To show that volume alone is an insufficient criterion, a solution is made for the characteristics of a traffic stream that satisfies the requirement of "the greatest good for the greatest number." The premise here is that maximum average speed is what the individual driver will find "good."

Thus, it is assumed that a segment of a continuous highway is performing best when it contains a number of vehicles traveling at such speed that the sum of the distance traveled per unit of time for each vehicle is maximized. Then in a length of highway, L, containing a number of vehicles, V, traveling at speed s, it is desired to maximize the product $V\ s$. But the number of vehicles is $V=L\ \rho$, in which ρ is the vehicle density, which is dependent on s.

* Generally, all V vehicles would not seek access to the highway at the same time, but so long as their rate of arrival equalled or exceeded q, Eq. A-1 would hold precisely.

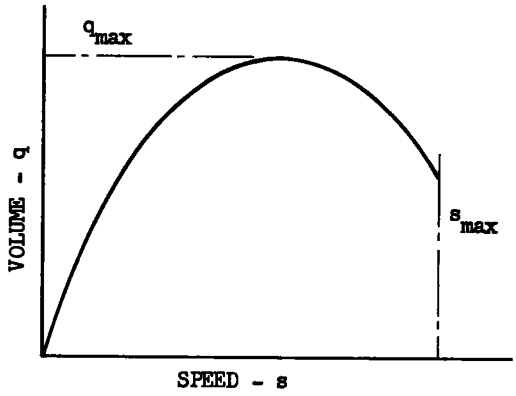


Figure A-1. Typical speed-volume relationship.

The maximum of the product $V\ s\ (= L\ \rho\ s)$ is found when the product of interdependent ρ and s is maximum. But $\rho\ s = q$, the rate of arrival of vehicles. Thus, the product $V\ s$ is a maximum at q_{max}, and volume *is* the measure of performance of traffic in one-dimensional flow. The commonly used unit would be vehicle-miles per hour, which has the dimensional unit of momentum when an average mass is assigned to each vehicle.

MATHEMATICAL DERIVATION OF THE COMPUTATIONAL LOGIC

The capacity of a particular segment of highway under constant environmental conditions and occupied by an average class of drivers can be described uniquely by a curve of speed, v, versus volume, q, or in terms of the parameter relating speed and volume, which is lane density, ρ. Inasmuch as it has been shown that maximum volume past a section of a continuous highway is the most efficient operating condition in terms of the amount of transportation taking place over the highway, it is desired to determine:

1. The maximum value of volume, q_{max}, that can be passed by the highway at a given time; and
2. The value of a descriptor of traffic behavior that will ensure that q_{max} can be obtained.

The value of lane density, ρ, would be a single criterion of what the maximum volume could be, regardless of what the volume was at any given instant in time. Long-term acquisition of speed and volume data will permit the generation of performance curves, which are, in fact, envelopes of maximum average speed-average volume points. Short-term performance at a section could be compared to the characteristic curve for that section to determine the level of performance.

DISCUSSION OF MEASURED TRAFFIC CHARACTERISTICS

Because the traffic detection equipment is unable to determine density in the lanes by direct observation, this quantity is computed by inference and extrapolation. For a detector over a particular lane, n, the total time, O_n, that the pavement directly under the detector is occupied during a data sampling period, τ, will be the sum of the pulses denoting "presence" multiplied by the time between pulses. Further, if the pulse return classification equipment discriminates between "high" and "low" vehicles, there will be a count of H_n commercial vehicles and L_n passenger vehicles during the interval τ. Assuming that the passenger vehicles have a statistically determined average length, $\bar{L}$, and that the similarly determined average length of a commercial vehicle is $\bar{H}$, the average speed, v_n, of the vehicles passing under the detector during the interval τ is expressed by

$$v_n = \frac{\bar{H} H_n + \bar{L} L_n}{O_n} \qquad \text{(A-3)}$$

However, for the purposes of determining the average performance for all lanes of the freeway at any station, x, the average speed for the time interval τ is given by

$$v_x = \frac{\bar{H} \Sigma H_n + \bar{L} \Sigma L_n}{\Sigma O_n} \qquad \text{(A-4)}$$

Average volume flow for the station during the time interval is given by

$$q_x = \frac{\Sigma H_n + \Sigma L_n}{\tau} \qquad \text{(A-5)}$$

and average lane density, ρ_r, for the region downstream of station x is computed from

$$\rho_x = \frac{q_x}{l_x v_x} \qquad \text{(A-6)}$$

in which l_x is the number of lanes in the freeway segment just downstream of station x.

It was previously stated that the average value of lane density given by Eq. A-6 is based on some inference and extrapolation. Because an average speed derived from a speed distribution has been determined for the point, the inference is drawn that the speed distribution will remain reasonably constant in the segment downstream of the station at which the measurements are made. It is realized that such an assumption is a zero order approximation, but once the computational logic is implemented it is expected that some empirically determined adjustments may have to be made so that correlation between computed and real performance will be satisfactory. The point density computed as in the foregoing is assumed to be extrapolated over a segment downstream of the measuring station equal to the product of the average speed, computed by Eq. A-4, and the data sampling period, τ. In general, it is expected that for the free-running traffic condition the sampling periods will be made compatible with the segment lengths between stations, d_x, so that

$$d_x \approx v_x \tau \qquad \text{(A-7)}$$

As values of q_x and ρ_x are computed for each station, x, and at successive time intervals, τ_i, they can be stored in the computer memory. An arbitrary number of values can be used as data points to construct a curve of q_x vs ρ_x for each station. It is proposed to construct the curves by fitting the equation of a parabola to the data by the method of least squares, and forcing the curve to pass through the origin. (It is obvious that when density is zero, volume must be zero.) To initiate the computations, an arbitrary parabola of reasonable coefficients can be put into storage for each station and the successive data points used to modify the curves. When the $(i+1)$th point is computed, the earliest point will be rejected from the computer's storage. Thus the stability of the curve will be a function of the number of successive points in time used for its solution.

Once the $(q\text{-}\rho)_x$ curves are brought to a current status they should remain essentially constant so long as the driver population and environmental conditions are unchanged. However, changing conditions will alter the curves so that expected performance will reflect the effects of a new environment (illumination, precipitation, visibility, roadway surface, etc.) Peak performance, q_{max}, that may be expected at any time can be determined by differentiating the equation of the parabola

$$q_x = D\rho_x + E\rho_x^2 \qquad \text{(A-8)}$$

with respect to ρ_x and setting $dq/d\rho = 0$. The solution yields the value of ρ at q_{max}, which may be called ρ_{crit}. The average speed that relates q_{max} and ρ_{crit} is determined from

$$v_{crit} = \frac{q_{max}}{\rho_{crit}\, l_x} \qquad \text{(A-9)}$$

CRITERIA FOR PREDICTION COMPUTATION

Figure A-2 shows that so long as average speed is maintained at or above v_{crit} the freeway will be in a free-running condition. That is, traffic will be free to regulate its movement so that the freeway will offer maximum service to all requiring this service. Should the average speed fall below v_{crit}, either of two possible causes may be the reason, as follows:

1. The drivers are not maintaining the maximum average speeds that the existing headways permit.
2. The freeway has become glutted to a density, $\rho > \rho_{crit}$, that forces the drivers to speeds lower than v_{crit} and volume below q_{max}.

Therefore, maintaining a current check on the status of density, ρ, provides an unambiguous indication of the status of traffic, insofar as its ability to use the freeway to greatest advantage is concerned.

Now, using current values of q, ρ, and v at each station, a logic is developed that will provide for the automatic determination of an impending traffic tie-up and a computation of travel time that may result if the tie-up is permitted to develop.

Let it be supposed that an incident of some sort has occurred in a freeway segment of length d_j, located just downstream of station j. One or more of the lanes become blocked and the previously freely-moving traffic is impeded by the bottleneck formed. The first parameter to be tested to determine conditions on the freeway is average speed. So long as speed remains at or above critical speed, the freeway is satisfying the needs of the drivers. But, should the situation arise in which $v_j < (v_{crit})_j$ it is necessary to examine further to see if this resulted from the drivers' election to dawdle or if density has forced the speed reduction. Thus, if

$$\rho_j \leq (\rho_{crit})_j \qquad \text{(A-10)}$$

no further action need be taken. But if

$$\rho_j > (\rho_{crit})_j \qquad \text{(A-11)}$$

this is a clear indication that a condition downstream of station j has forced the previously free-running traffic to compress to a point where volumes less than q_{max} will be realized—that is, congestion has set in.

Thus, the existence of a self-limiting traffic flow condition can be determined by automatic equipment if the following inequalities are computed at any time and station:

$$v_j < (v_{crit})_j \text{ and } \rho_j > (\rho_{crit})_j \qquad \text{(A-12)}$$

From the q-ρ curve for a station operating at the conditions given in Eq. A-12 can be determined the maximum flow, $(q)_{\rho_j}$, that the bottleneck will pass for the current value of ρ_j. Whether or not the value of $(q)_{\rho_j}$ will force a queue to form will depend on traffic conditions upstream of station j. If the oncoming volumes are less than or equal to $(q)_{\rho_j}$, there should be little if any delay involved in getting the traffic to move through the bottleneck area.

Similarly, should an interrogation of the computer memory show that the first current value of volume in excess of $(q)_{\rho_j}$ is well upstream of station j, this should be sufficient reason to defer a computation of predicted travel time. Because the automated system will not be able to determine the reason for the existence of the conditions expressed in Eq. A-12 nor how long they will persist, the sounding of an alarm at this time might be too conservative. The road might recover before the excess volume moved downstream to the bottleneck station. Therefore, a first trial logic programmed into the computer must satisfy the additional condition that

$$q_{j+1} > q_j \qquad \text{(A-13)}$$

at any time, t, before a computation of predicted travel time will be triggered.

(*Note:*—A comment is offered here concerning the use of both speed and density in the conditional inequalities of Eq. A-12. It has been stated that lane density is a sufficient criterion for expected performance maximums at any station. Remembering that the q-ρ curves from which these performance values are read were constructed on a statistical basis, it is natural to expect that the distribution will contain points of higher performance. Thus, a unilateral decision that the freeway operation is deteriorating, due to a reading of density above ρ_{crit}, is avoided (no false alarm) if it is found that average speeds are continuing at or above v_{crit}.)

FORMATION OF THE QUEUE

Once the decision has been made to make a prediction of travel time (Eqs. A-12 and A-13 are satisfied), the problem becomes one of estimating travel time of vehicles that are forecast to make up the queue that will form. Generally, this entails an assessment of the traffic upstream of the bottleneck in order to determine those volumes that exceed the bottleneck flow, and an estimate of the time that will elapse before the last vehicles that will form the queue will be forced to reduce speed and headway as they enter the queue. The computations may be complicated by the fact that there can be one or more stations in the upstream region at which the flow is less than bottleneck volume, in an otherwise more-than-bottleneck-flow stream. Although this situation may not occur over the relatively limited length of the Proving Ground on the John C. Lodge Freeway, the logic is written to handle such a contingency in order to make it more generally applicable.

In the original statement of the mathematical model, the computational logic sought the last station upstream of the bottleneck for which the volume, q_k, was greater than the bottleneck volume, q_j. Then, successive stations upstream of station k were investigated, iteratively, to determine when the time of passage through the bottleneck was less than the free-running travel time, based on the volume at station x and the maximum speed for that volume, as read from the performance envelopes for each segment.

In the process of computation, an interception geometry was worked out for the position of vehicles, originally at station k, at the time they compressed into the queue. It turns out that it was unnecessary and probably not accurate. It is superseded by simply using the iterative procedure written for stations upstream of station k.

But first a theoretical approach is shown for location of the position of the bottleneck, assuming one-dimensional flow conditions in a segment. If the data indicate that it may be practical to implement this logic, it replaces the assumption of the location of the bottleneck at $d_j/2$ downstream of station j. Otherwise, the assumption is retained.

Location of Bottleneck, One-Dimensional Flow

Let it be assumed that, at any time, t, a bottleneck occurs in a steady traffic stream of volume q, which is moving at velocity v (see Fig. A-2). Then, at some time or times later, the volumes at stations j and j − 1 will decrease. The first point of interest is to determine when q actually falls off at the two stations.

If traffic were truly a stream, the change could be detected easily. The volume rate could be sensed instantaneously from pitot and absolute pressure readings (for a compressible flow). Individual time headways might be used as an indication of instantaneous volume reduction, but it is suspected that the variation in arrival rates in a uniform traffic stream might exceed the average change due to the change in volume. But neither of these methods is being used to measure volume; it is computed from Eq. A-5 and will have variations over successive τ, even for steady long-term flows ($\tau \approx 20$ to 60 sec; long-term ≈ 5 min).

The problem of precise location of the bottleneck is examined theoretically. Then, if a satisfactory solution is found, further investigation can be made to determine its practical validity.

The bottleneck can occur at any time during the period τ, (and at any place over the length d_j). Let it be assumed that a reduction in volume occurs during τ. Typically, it may look like Figure A-3. However, if a steady flow is postulated, Case c is ruled out. Actually, Case b becomes a special version of Case a ($c = 0$). Then,

$$q_\tau = \frac{c\, q_{-\tau} + (1 - c)\, q_{+\tau}}{\tau} \qquad \text{(A-14)}$$

from which

$$c = \frac{q_\tau - q_{+\tau}}{q_{-\tau} - q_{+\tau}} \qquad \text{(A-15)}$$

Values of c may be determined, in this manner, for both stations j and j — 1, even though the step function may have occurred during different τ.

Now, having two times at which volume reduction signals reached stations j and j — 1, it is possible to determine the distance, a, from station j to the bottleneck. If t_j and t_{j-1} are the times at which the signals reached the stations and t_a is the time at which the incident occurred,

$$t_{j-1} - t_a = \frac{d_j - a}{v} \qquad \text{(A-16)}$$

and

$$t_j - t_a = a / v_s \qquad \text{(A-17)}$$

in which v_s is the shock wave propagation speed and v is the speed of the traffic stream during the period $-\tau$, as measured at station j. By simultaneous solution,

$$a = \left(\frac{v\, v_s}{v + v_s}\right)\left(t_j - t_{j-1} + \frac{d_j}{v}\right) \qquad \text{(A-18)}$$

Pipes * develops the expression for the speed of the shock wave, which is written in the present notation as

$$v_s = \frac{q_{-\tau} - q_{+\tau}}{\rho_{-\tau} - \rho_{+\tau}} \qquad \text{(A-19)}$$

An obvious complication in the computational procedure is injected by the use of this entire bottleneck location routine. The prediction must be delayed for at least the interval τ, in order to obtain the values $q_{+\tau}$ and $\rho_{+\tau}$. Also, if the bottleneck occurred very close to station j and if $d_j \gg v\,\tau$, the search for a reduction in volume at station j — 1 might be protracted to longer than the interval $+\tau$.

Thus, the first computation will be to estimate the number of vehicles contained between the bottleneck and station j. Because the first intimation of a slow-down in segment d_j will be the conditions expressed by Eqs. A-12 and A-13, there is no choice, at time t, other than to make the prediction based on q_j. The situation may change at time, $t + \tau$, but that would be a matter that would call for repeating this whole process. As previously suggested, the number of vehicles is computed from

$$N_j = a\, d_j (\rho)_{q_j} \qquad \text{(A-20)}$$

in which $(\rho)_{q_j}$ is the maximum expected value of density for the volume, q_j, read from the q-ρ curve for station j.

* "An Introduction to Traffic Flow Theory." *HRB Spec. Report 79*, pp. 3-5 (1964).

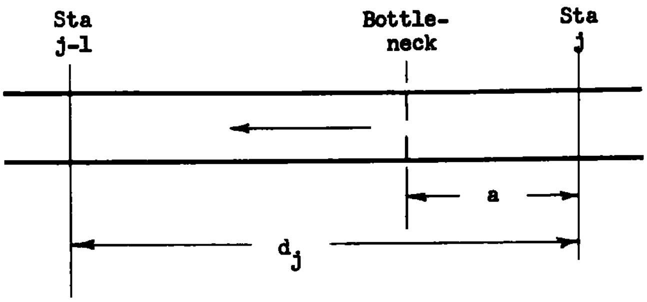


Figure A-2. Geometry of bottleneck occurrence.

In theory, the growth rate of the queue would be given by

$$\frac{dQ}{dt} = \frac{q - q_j}{l(\rho)_{q_j}} \qquad \text{(A-21)}$$

in which

q = volume rate approaching the tail of the queue;
q_j = bottleneck volume;
l = number of lanes at the tail end of the queue;
$(\rho)_{q_j}$ = lane density for the bottleneck volume; and
Q = length of the queue.

Unfortunately, the variability of the traffic stream makes it impractical to consider the use of the differential form to solve for the length of the queue when the numerator of the right-hand side, $q - q_j$, has gone to zero. Instead, the solution is found by an iterative process, which it is hoped will be satisfactorily accurate, particularly when it has been modified in the light of experience.

The numbers of vehicles in the segments upstream of station j successively are computed. It has been suggested that the frequency of the detector stations should be such that

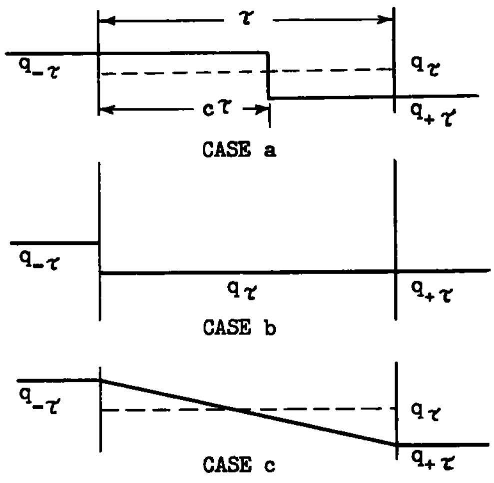


Figure A-3.

$$d_x \approx v_x \tau \qquad \text{(A-22)}$$

and if this approximation was an equality, the number of vehicles in any segment (d_x, for instance) would be

$$N_x = \tau\, q_x \qquad \text{(A-23)}$$

Ordinarily, $d \neq v\, \tau$ for any segment, so it is suggested that the stored data for past time intervals be consulted in order to compute N_x. At any station, x, the average distance traveled by the vehicles during τ will be $v_x \tau$. Then, at time t, if $v_x\, \tau \geqq d_x$ it may be assumed that

$$N_x = \left(\frac{dq}{v}\right)_x \qquad \text{(A-24)}$$

(Any spillage from d_x into d_{x-1}, the downstream segment, may be ignored, because it is not expected that $v\, \tau > d$ by any considerable margin, for most segments.) However, if $v_x\, \tau < d_x$, it is appropriate to include the values of speed and volume prior to time t. This will be done by summing successively earlier values of $v_x\, \tau$ until

$$(v_x\, \tau)_t + \ldots + (v_x\, \tau)_{t-p\,\tau} \geqq d_x \qquad \text{(A-25)}$$

in which p takes on increasing integral values, beginning with unity, until the condition of Eq. A-25 is satisfied. The number of vehicles in segment d_x is computed from

$$N_x = \left[\tau \sum_{p=1}^{p=p} (q_{t-(p-1)\tau})_x\right] + \left[\frac{d_x - \tau(v_t + \ldots v_{t-(p-1)\tau})_x}{(v_{t-p\tau})_x}\right][(\tau q_{t-p\tau})x] \qquad \text{(A-26)}$$

The value of p in Eq. A-26 will have been determined in Eq. A-25.

Theoretically, if speeds on the freeway become very low, it is possible for p to become very large. The requirements for computer storage space could become unrealistic and the control system could be hampered by the computing time required. Therefore, an arbitrary upper limit is placed on the value of p, and if the condition of Eq. A-25 is not satisfied by the time this upper limit, i, is reached, an average value of density during i_T will be used to determine N_x at time t. The equation has the general form of Eq. A-24; that is,

$$N_x = \left[\frac{d(q_t + \ldots + q_{t-i\tau})}{v_t + \ldots + v_{t-i\tau}}\right]_x \qquad \text{(A-27)}$$

The physical implication of Eq. A-27 is that traffic has slowed and compressed to the degree that a fairly uniform density might be expected over the length of the segment and that an average of the immediate past may be adequate to represent the whole content of the segment when extrapolated downstream.

DISSIPATION OF THE QUEUE

Having developed the expression for the numbers of vehicles in the freeway segments that may contribute to the formation of a queue, each segment upstream of station j is considered in turn. Travel times to pass from stations j + 1, j + 2, . . . etc., are computed based on the aggregate numbers of vehicles between the station and the bottleneck, and the bottleneck volume. It is assumed that the bottleneck volume, q_j, remains constant over the prediction period. Thus, the minimum possible time for a vehicle presently at station x to intercept the bottleneck is

$$(t_i)_x = \frac{N_j + N_{j+1} + \ldots + N_x}{q_j} \qquad \text{(A-28)}$$

The normal travel time from station x to the point at which the bottleneck is located, $(t_n)_x$; in the absence of the queue, could be as short as that obtained by assuming that the vehicle at station x passes through successive segments at an average "maximum performance" speed, v_{ave}. Thus,

$$(t_n)_x = \frac{a\, d_j + d_{j+1} + \ldots + d_{x-1}}{(v_{ave})_x} + \frac{d_x}{v_x} \qquad \text{(A-29)}$$

Then, starting with station j + 1, as the computations of Eqs. A-28 and A-29 are made the values are compared until some

$$(t_n)_x \geqq (t_i)_x \qquad \text{(A-30)}$$

The vehicles at that station x should be the last ones affected by the bottleneck.

COMPUTATION OF AVERAGE SPEED

The value of predicted average speed, $(v_{ave})_x$, used in Eq. A-29 is believed to be the most elusive quantity in this entire procedure. As before, the assumption made here regarding the speed of any platoon of drivers, N_x, in a segment d_x, is that they will move forward at the maximum average speed consistent with their volume demand, q_x, and the ability of any section of freeway to pass the flow. Only one restriction is placed on this assumption; it is suggested that passing from one platoon into another will not be considered, therefore no average speed computation for a following platoon will be permitted to exceed that of a downstream platoon.

As each station x is considered in turn, the speed over the segment downstream will be the current v_x. For the segments further downstream, the value of q_x will be used to find the maximum speed of passage for this volume for each station. These speeds, $(v_q)_x$, will be read from the left-hand portion of the appropriate q-ρ curves (Fig. A-2) The form of the expression for average speed would be:

$$(v_{ave})_x = \frac{\frac{d_j}{2}(v_q)_j + \sum_{j+1}^{x-1} (dv_q)_x + d_x\, v_x}{\frac{d_j}{2} + \sum_{j+1}^{x-1} (d)_x + d_x} \qquad \text{(A-31)}$$

If at any station, call it z, $q_x > (q_{max})_z$, *the speed* $(v_q)_x$ would be determined by the delay forced on a vehicle currently at the upstream end of segment x. Speed over the segment, d_z, would be found from

$$(v_q)_z = \frac{(v_{crit}\, q_{max})_z}{q_x} \qquad \text{(A-32)}$$

Finally, if the average speed found from Eq. A-31 was found to exceed the average speed of the vehicles from segment x-1, the latter value would be used. Thus,

$$(v_{ave})_x \leq (v_{ave})_{x-1} \qquad \text{(A-33)}$$

OUTLINE OF THE COMPUTATIONAL PROCEDURE

The following outlines the computational logic to be used in order to predict that time in the future when a queue that is threatening to form as the result of a present traffic blockage on a freeway will be dissipated.

1. Compute traffic performance characteristics at each data collection station, x, for successive sampling periods.
   (a) Compute average vehicle speed, v_x, for the data sampling period, τ, by the use of Eq. A-4. Values for ΣH_n, ΣL_n, and ΣO_n will have been computed and stored. $\overline{H}$ and $\overline{L}$ are stored input values.
   (b) Store the values of v_r computed in 1(a) for i sampling periods, after which reject the earliest value for v_x when a new one is computed. The value of i will be an input.
   (c) Compute average vehicle passage rate, q_x, for the data sampling period by the use of Eq. A-5. The value for τ is a stored input.
   (d) Store the values of q_r computed in 1(c) for i sampling periods, after which reject the earliest value for q_r when a new one is computed.
   (e) Compute average lane density, ρ_x, for the data sampling period by the use of Eq. A-6. The value of l_x will be a stored input.
   (f) Store the values of ρ_x computed in 1(e) for i sampling periods, after which reject the earliest value for ρ_r when a new one is computed.
2. Compute traffic capacity curves at each data collection station, x, for successive sampling periods.
   (a) At the end of each data sampling period, τ, solve for the coefficients D and E for a curve of the form:

$$q_r = D\rho_x + E\,\rho_x^2 \qquad \text{(A-8)}$$

   Arbitrary values of D and E will be supplied as stored inputs, after which the computed values of q_x and ρ_r, up to i ordered pairs, will be used in a least squares fit routine to modify the values of D and E.
   (b) Store the current values for D and E for each station.
   (c) Solve for maximum volume, $(q_{max})_x$ by differentiating Eq. A-8 and maximizing. Store the current value of $(q_{max})_x$.
   (d) Compute critical density, $(\rho_{crit})_x$, by using the value obtained in 2(c) in Eq. A-8. Store the current value of $(\rho_{crit})_x$.
   (e) Compute critical speed, $(v_{crit})_x$, by use of Eq. A-9 and values from 2(c) and 2(d). Store the current value of $(v_{crit})_x$.
3. Determine if conditions exist at any station that warrant the exercising of the travel time prediction logic.
   (a) Test if any current $v_x < (v_{crit})_x$. If no, return to instruction 1 at next sampling period. If yes, continue below.
   (b) For each station satisfying the condition of 3(a), test if $\rho_x > (\rho_{crit})_x$. If no, return to instruction 1 at next sampling period. If yes, continue below.
   (c) For each station satisfying the condition of 3(b), test if $q_x < q_{x+1}$, where x + 1 is the station upstream of any station x satisfying the conditions of 3(a) and 3(b). If no, return to instruction 1 at next sampling period. If yes, designate each such station as station j and continue below with each such station separately.
4. Compute minimum travel time from stations x upstream of station j.
   (a) Compute number of vehicles, N_j, downstream of station j, using Eq. A-20.
   (b) All computations from this point on shall be accomplished serially for each station x, beginning with x = j + 1.
   (c) For each station x (j + 1, j + 2, . . . etc.), compute the value of $v_x\tau$ and store. Compare each value of $v_x\tau$ with those of d_x, which are stored inputs. If $v_x\tau \geq d_x$, compute the number of vehicles, N_x, in each segment, d_x, from Eq. A-24 and store. If $v_x\tau < d_x$, continue below.
   (d) Designate the current $v_x\tau$ as $(v_x\tau)_t$. Then, adding successively earlier values of $v_x\tau$, continue until the condition of Eq. A-25 is satisfied, or $p = i$. If Eq. A-25 is satisfied with $p \leq i$, compute N_x from Eq. A-26 and store. If Eq. A-25 is not satisfied with $p = i$, compute N_x from Eq. A-27 and store.
   (e) Compute minimum possible travel time from station x to the bottleneck, $(t_t)_x$, using Eq. A-28 and store. Values for N_j, N_{j-1}, etc. will be from instructions 4(a) and 4(c) or 4(d).
5. Compute predicted average speed of vehicles closing on the queue from segments upstream of station j.
   (a) Determine successively, q_r for each station upstream of station j from instruction 1(c).
   (b) From instruction 2(c), determine q_{max} for each station downstream of station x. Test if any $q_{max} < q_r$, the latter from 5(a).
      (1) If yes, determine v_{crit} for each such station from instruction 2(e). Compute speed over the segment downstream of each station from

$$v = \frac{v_{crit}\, q_{max}}{q_x} \qquad \text{(A-32a)}$$

      and store.
      (2) If no, continue below.
   (c) Determine average speed over the segment downstream of each station.
      (1) Compute lower value of ρ at q_x, using Eq. A-8 and D and E from instruction 2(a).
      (2) Compute each

$$v = \frac{q_r}{\rho\, l_x} \qquad \text{(A-9}a\text{)}$$

   (d) Compute average speed using an equation of the form of Eq. A-31. Initially it will be $(v_{ave})_{j+1}$. Values of the speeds for the individual segments, v, are from 5(b)(1) and 5(c)(2) and segment lengths, d, are from stored inputs.
   (e) Test if the average speed from instruction 5(d), $(v_{ave})_x > v_{x-1}$. If yes, let $(v_{ave})_x = v_{x-1}$ and continue at 6(a). If no, use $(v_{ave})_x$ and continue at 6(a).

6. Compute normal travel time and compare with minimum possible travel time under the bottleneck conditions.
   (a) Compute a normal travel time, $(t_n)_x$, by the use of an equation of the form of Eq. A-29.
   (b) Test if $(t_n)_x \geqq (t_i)_x$. The value of $(t_i)_x$ is from instruction 4(e). If no, store $(t_i)_x$ and return to instruction 4(b). If yes, print out $(t_n)_x$ and the identification of station x as that at which the vehicles are predicted not to suffer any delay. END COMPUTATION.

2. TRAVEL TIME AS A MEASURE OF ROAD NETWORK EFFICIENCY AND THE ACQUISITION OF EFFICIENCY ELEMENTS

Traffic engineers have used vehicular volume flow as a measure of road performance and capacity for almost as long as the profession has existed. Although there is little denial that it is an appropriate criterion for determining requirements for new construction and for evaluating the ability of a road system to comply with a demand, volume flow does not reflect directly the ability of a road system to furnish service to the individual driver, whose interest is centered more nearly in the time required to traverse the route of travel.

In the following discussion two values of travel time are used as a measure of road network performance. One is actual travel time experienced in the real system and the other is travel time over the same route in the absence of all hindering traffic and control elements. The networks of Figure A-4 are considered, Figure A-4a being a main city street with its cross streets, and Figure A-4b being a limited-access road with some of its entrances and exits.

The obvious reasons why a vehicle finds itself in such a network are as follows:

1. To use the network as a means of passage from an origin outside the network to a destination (a) inside the network, or (b) outside the network.
2. To start from a point within the network and to terminate the trip (a) within the network, or (b) outside the network.

This work is concerned only with vehicles whose origins and destinations are outside the network.

EFFICIENCY

Consider that there are v_i vehicles leaving the i^{th} exit per minute, say, on the average (generally uniformly distributed over the interval).

The number of vehicles leaving thru exit i over the time interval of 1 min is thus

$$N_i = V_i\, 1 \tag{A-34}$$

These vehicles, however, entered from entrances 1, 2, 3, . . . k < i, so

$$N_i = N_{i1} + N_{i2} + \ldots + N_{ik} \tag{A-35}$$

where N_{ij} = the number of vehicles that leave by exit i that entered at entrance j, or the number of vehicles leaving the network through exit j consists of the sum of those leaving that entered at entrance 1, 2, . . . , k.

Associated with each N_{ij} is the distance from entrance j to exit i, or S_{ij}. Let T_{ij}' denote the time of traverse of distance S_{ij} in the absence of any other traffic; and T_{ij} the actual time of traverse of S_{ij} in the presence of traffic. Hence, the total distance traveled by the N_i vehicles is

$$S_i = N_{i1}\, S_{i1} + N_{i2}\, S_{i2} + \ldots + N_{ik}\, S_{ik} \tag{A-36}$$

the cumulated time in making the trips is

$$T_i = N_{i1}\; T_{i1} + N_{i2}\; T_{i2} + \;\ldots\; + N_{ik}\; T_{ik} \tag{A-37}$$

and clearly the minimum time in making the trips is

$$T_i' = N_{i1}\, T_{i1}' + N_{i2}\, T_{i2}' + \ldots + N_{ik}\, T_{ik}' \tag{A-38}$$

An index of interaction delay due to traffic is given by

$$I_i = \frac{T_i}{T_i'} = \frac{\sum_{j=1}^{k} N_{ij}\, T_{ij}}{\sum_{j=1}^{k} N_{ij}\, T_{ij}'} \tag{A-39}$$

and one might say that "the trips were I_i times as bad as they would have been in the absence of interacting traffic." Indeed, a measure of the efficiency of the traffic flow to exit i is arrived at by merely inverting Eq. A-39 to obtain

$$E_i = \frac{T_i'}{T_i} \tag{A-40}$$

where 100% efficiency means that each vehicle went from its origin to its destination with no interaction with other vehicles.

ELEMENTARY EFFICIENCIES

An elementary efficiency is defined as the efficiency from a single entrance to a single exit. Hence,

$$\begin{aligned}
E_{i1} &= \frac{N_{i1}\, T_{i1}'}{N_{i1}\, T_{i1}} = \text{efficiency from entrance 1 to exit i} = \frac{T_{i1}'}{T_{i1}} \\
E_{i2} &= \frac{N_{i2}\, T_{i2}'}{N_{i2}\, T_{i2}} = \text{efficiency from entrance 2 to exit i} = \frac{T_{i2}'}{T_{i2}} \\
&\;\;\vdots \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \vdots \\
E_{ik} &= \frac{N_{ik}\, T_{ik}'}{N_{ik}\, T_{ik}} = \text{efficiency from entrance k to exit i} = \frac{T_{ik}'}{T_{ik}}
\end{aligned} \tag{A-41}$$

Then Eq. A-40 can be written in terms of the elementary efficiencies, or

$$E_i = \cfrac{1}{\cfrac{1}{E_{i1}}\cfrac{N_{i1}T_{i1}'}{\Sigma N_{ij}T_{ij}'} + \cfrac{1}{E_{i2}}\cfrac{N_{i2}T_{i2}'}{\Sigma N_{ij}T_{ij}'} + \ldots + \cfrac{1}{E_{ik}}\cfrac{N_{ik}T_{ik}'}{\Sigma N_{ij}T_{ij}'}} \tag{A-42}$$

where it is recognized that

$$P_{i1} \equiv \frac{N_{i1}T_{i1}'}{\Sigma N_{ij}T_{ij}'} \tag{A-43}$$

which is the percentage of i-time contributed by vehicles entering through entrance 1 under ideal conditions. It also should be noted that

$$\sum_{j=1}^{k} P_{ij} = 1 \tag{A-44}$$

Hence, the elementary efficiencies behave as resistors in a parallel network, and

$$\frac{1}{E_i} = \frac{P_{i1}}{E_{i1}} + \frac{P_{i2}}{E_{i2}} + \ldots + \frac{P_{ik}}{E_{ik}} \tag{A-45}$$

as shown in Figure A-5.

ELEMENTARY INDICES

Of course, the foregoing may be written in terms of elementary indices. By letting $E_i = \frac{1}{I_i}$ and $E_{ij} = \frac{1}{I_{ij}}$, where $j = 1, 2, \ldots, k$,

$$I_i = \frac{1}{E_i} = P_{i1}I_{i1} + P_{i2}I_{i2} + \ldots + P_{ik}I_{ik} \tag{A-46}$$

The indices then assume the series resistor schematic depicted in Figure A-6.

EXTENSION TO ALL M EXITS

It should be recalled that the efficiency E_i developed in the foregoing is the efficiency of vehicles leaving exit i during the course of a single minute. In the following the result is extended to all m exits.

The efficiencies at exits 1, 2, . . . , m, are, respectively,

$$\left.\begin{aligned} E_1 &= \frac{T_1'}{T_1} \\ E_2 &= \frac{T_2'}{T_2} \\ &\vdots \\ E_m &= \frac{T_m'}{T_m} \end{aligned}\right\} \tag{A-47}$$

and the over-all efficiency of the network is

$$E = \frac{T_1' + T_2' + \ldots + T_m'}{T_1 + T_2 + \ldots + T_m}$$

$$= \cfrac{1}{\cfrac{1}{E_1}\cfrac{T_1'}{\sum_{i=1}^{m} T_i'} + \cfrac{1}{E_2}\cfrac{T_2'}{\sum_{i=1}^{m} T_i'} + \ldots + \cfrac{1}{E_m}\cfrac{T_m'}{\sum_{i=1}^{m} T_i'}} \tag{A-48}$$

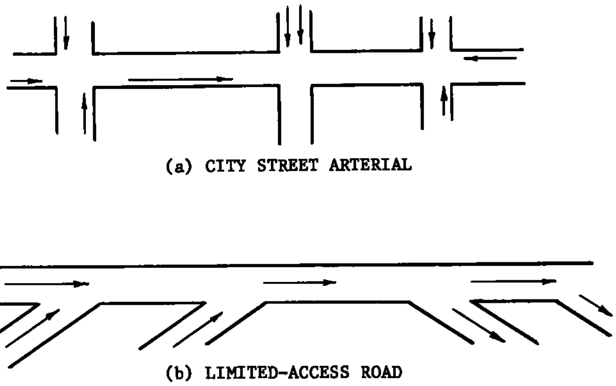


Figure A-4. Typical subnetworks for traffic flow analysis.

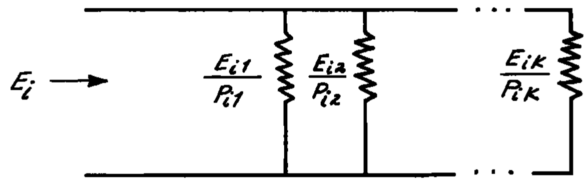


Figure A-5. Efficiency at exit i in terms of elementary efficiencies.

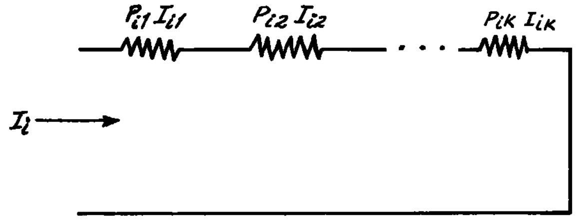


Figure A-6. Index at exit i in terms of elementary indices.

where

$$P_1 \equiv \frac{T_1'}{\sum_{i=1}^{m} T_i'} \tag{A-49}$$

which is the percentage of total time contributed by vehicles leaving through exit 1, under ideal conditions, and

$$\sum_{i=1}^{m} P_i = 1 \tag{A-50}$$

Here, of course,

$$\sum_{i=1}^{m} T_i' = \sum_{i=1}^{m}\sum_{j=1}^{k} N_{ij}T_{ij}' \tag{A-51}$$

and once again writing the index $I = 1/E$, there result the network equivalents shown in Figure A-7.

The purpose of this discussion is to define elements that currently can be measured without disturbing the traffic stream to enable calculation of the efficiency of a road network. Using the form of Eq. A-40 (that is, $E = T'/T$), consider flow through a traffic network between, say, the hours of 8:00 and 8:15 AM. Let it be assumed that there are k entrances and m exits; that e_j vehicles enter through entrance j; $j = 1, 2, 3, \ldots, k$; and

$$\sum_{j=1}^{k} e_j = \zeta \tag{A-52}$$

Also, that l_i vehicles leave through exit i; $i = 1, 2, 3, \ldots, m$; and

$$\sum_{i=1}^{m} l_i = L \tag{A-53}$$

Then the entrance percentages are

$$\frac{e_1}{\zeta}, \frac{e_2}{\zeta}, \ldots, \frac{e_k}{\zeta} \tag{A-54}$$

where e_j/ζ is the percentage of entering vehicles that entered via entrance j, and the exiting percentages are

$$\frac{l_1}{L}, \frac{l_2}{L}, \ldots \frac{l_m}{L} \tag{A-55}$$

where l_1 is the percentage of leaving vehicles which left by exit i.

Then the entrance/exit percentage matrix (ζ/L matrix) is

(A-56)

Entrance		1	2	3	k	$\sum_{j=1}^{k}$
Exit	1	P_{11}	P_{12}	P_{13}	P_{1k}	l_1/L
	2	P_{21}	P_{22}	P_{23}	P_{2k}	l_2/L
	3	P_{31}	P_{32}	P_{33}	P_{3k}	l_3/L
	:	:	:	:	:	
	m	P_{m1}	P_{m2}	P_{m3}	P_{mk}	l_m/L
$\sum_{i=1}^{m}$		e_1/ζ	e_2/ζ	e_3/ζ	e_k/ζ	1

where the entry

$$P_{ij} = \frac{l_i}{L}\frac{e_j}{\zeta} \tag{A-57}$$

bespeaks a partition of the "fluid in the pipe" in accordance with network measurement and a most elementary model. It suffers from the deficiencies of the macroscopic view, for it considers input flow and output flow to be statistically independent. It does not attempt to account for time of transit, but does give a mathematically tight representation of the facts at hand. Clearly, it can be supplanted by integrating minutiae of an exact individual vehicle origin-destination study at great expense; or it can be more closely approximated by enhancement through sampling studies at somewhat less cost. For the present, the set of P_{ij} is accepted as a (crude) measure of present road usage and the travel times are weighted accordingly.

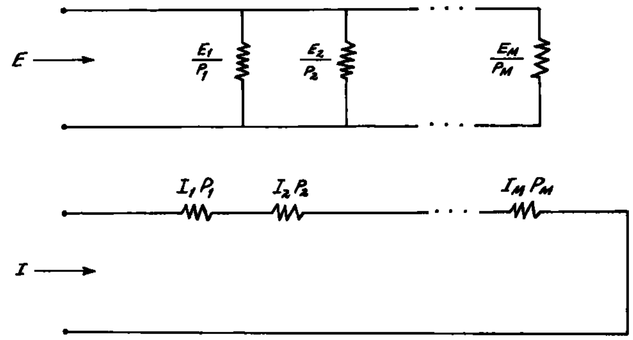


Figure A-7. Parallel representation of efficiency (upper) and series representation of index (lower).

Consider the corresponding matrix of ideal times (in the absence of traffic)

$$\begin{pmatrix} T_{11}' & T_{12}' & T_{13}' & T_{1k}' \\ T_{21}' & T_{22}' & T_{23}' & T_{2k}' \\ \vdots & & & \\ T_{m1}' & T_{m2}' & T_{m3}' & T_{mk}' \end{pmatrix} \tag{A-58}$$

and that of actual times

$$\begin{pmatrix} T_{11} & T_{12} & T_{13} & T_{1k} \\ \vdots & & & \\ T_{m1} & T_{m2} & T_{m3} & T_{mk} \end{pmatrix} \tag{A-59}$$

where T_{ij}' is the ideal time from entrance j to exit i, in the absence of traffic, and T_{ij} is the actual time from entrance j to exit i. The efficiency is then given by

$$E = \frac{E(T')}{E(T)} \tag{A-60}$$

where

$$E(T') = \sum_{j=1}^{k} \sum_{i=1}^{m} P_{ij} T_{ij}' \tag{A-61}$$

and

$$E(T) = \sum_{j=1}^{k} \sum_{i=1}^{m} P_{ij} T_{ij} \tag{A-62}$$

MEASUREMENT OF EFFICIENCY ELEMENTS

From the foregoing it is clear that road network efficiency can be calculated with a knowledge of the following:

1. The origins and destinations of each vehicle in the network.
2. The time of travel of each vehicle from origin to destination, including all queuing delays.
3. The time of travel of each origin-destination route, in the absence of traffic (ideal time).

However, it is also obvious that such information cannot readily be collected without an inordinate disturbance of the traffic stream.

The calculation of efficiency thus consists of three measurable factors, as follows:

1. P_{ij} — This is determined from input and output measurements of volume.
2. T_{ij}' — This is an ideal time and can be readily evaluated from the physical (road, traffic signals, signals, stop sign) characteristics of the road network, running the course.
3. T_{ij} — Average car in a traffic stream from entrance j to exit i.

It will be noted that maximizing the efficiency, E, for a particular road is equivalent to minimizing actual time $E(T)$. That the concepts of efficiency and travel time are essentially different, however, is illustrated by the following example.

EXAMPLE: Let a network be assumed of efficiency $E = \frac{E(T')}{E(T)}$. Also, let an improvement be contemplated in the network such that the new expected actual travel time is $E(T) - \Delta$ and the new ideal travel time is $E(T') - \delta$. Then the new efficiency is

$$E' = \frac{E(T') - \delta}{E(T) - \Delta} \tag{A-63}$$

It is desired to determine if the efficiency has improved. This is tested by the following comparisons:

$$E' > E \tag{A-64}$$

or

$$\frac{E(T') - \delta}{E(T) - \Delta} > \frac{E(T')}{E(T)} \tag{A-65}$$

or

$$\frac{E(T)}{E(T')} [E(T') - \delta] > E(T) - \Delta \tag{A-66}$$

or

$$\Delta > \frac{E(T)}{E(T')} \delta \tag{A-67}$$

or

$$\frac{\delta}{\Delta} < E \tag{A-68}$$

The new efficiency is higher than the original efficiency when (Eq. A-68) the ratio of the improvement in ideal travel time to the improvement in actual travel time is less than the original efficiency. This last expression is somewhat intuitive.

CONCLUDING COMMENTS

1. The technique discussed here may be applied to determining that point in a traffic network whose improvement would contribute most to improvement of over-all efficiency; in other words, to call attention to "public enemy number one."

2. The technique may also be applied to measure the degree to which a road improvement is actually being utilized and thus point to the necessity for enhanced driver education, communication by road signs, and the like.

3. The matrices given (Eqs. A-56, A-58, and A-59) assume that all exits are fed by all entrances. This, of course, requires modification depending on the actual arrangement.

4. The over-all delay time, D, is readily calculable from the efficiency; that is,

$$E = \frac{T'}{T} = \frac{T'}{T' + D} \tag{A-69}$$

and the percentage of traffic delay is

$$\frac{D}{T'} = \frac{1 - E}{E} \tag{A-70}$$

It should be noted, also, that the traffic delay exceeds 100% when the efficiency is less than 50% ($E < 0.50$).

3. IMPLICATIONS OF A LINEAR SPEED-DENSITY RELATIONSHIP

This section examines the implications of a linear relationship between speed and density on the following traffic parameters:

1. Volume.
2. Spacing.
3. Transit time.
4. Time headway.

It also shows that a linear speed-density relationship implies that:

1. The characteristics of speed-volume and density-volume are parabolic in form.

2. Maximum volume occurs when vehicle spacing is two car lengths at a speed of one-half maximum speed, assuming maximum density is "bumper-to-bumper."

3. Time headway is improperly utilized, because time headway minimizes at maximum volume and increases at both lower and higher speeds.

NOTATION

In this section the following symbols are used:

u = mean vehicle speed;
u_0 = vehicle speed at zero density;
ρ = traffic density;
ρ_0 = traffic density at zero speed, "bumper-to-bumper";
V = volume = number of vehicles per hour;
V_m = maximum volume;
u_m = vehicle speed at maximum volume;
ρ_m = traffic density at maximum volume;

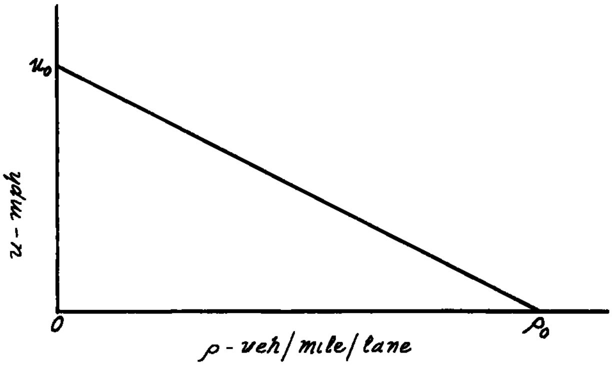


Figure A-8. Speed vs density, on linear assumption.

$\delta =$ vehicle spacing;
$\delta_0 =$ vehicle spacing at zero speed;
$\delta_m =$ vehicle spacing at maximum volume;
$T =$ transit time over a fixed distance, S;
$T_m =$ transit time at speed u_m;
$\tau =$ time headway.

LINEAR ASSUMPTION

It is assumed that traffic speed, u miles per hour, is regulated in a linear fashion by traffic density, ρ vehicles per mile per lane, as shown in Figure A-8.

The relative speed, u/u_0, and the relative density, ρ/ρ_0, satisfy

$$\frac{u}{u_0}+\frac{\rho}{\rho_0}=1 \qquad \text{(A-71)}$$

where u_0 is the maximum speed, attained at zero density, and ρ_0 is the maximum density, attained at zero speed. This equation shows that a speed, u, of K % of its maximum value, u_0, implies a density, ρ, of $(100-K)$% of its maximum value, ρ_0. For example, a relative speed of 40% implies a relative density of 60%.

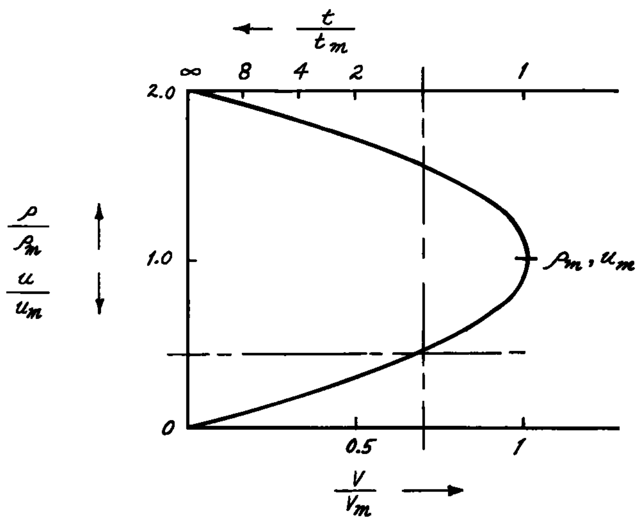


Figure A-9.

VOLUME

The volume, V, is the number of vehicles passing a given point in one hour. Hence,

$$V=\rho\, u \text{ vph} \qquad \text{(A-72)}$$

However, from Figure A-8,

$$V(\rho_o)=V(u_o)=0 \qquad \text{(A-73)}$$

Volume can be expressed solely in terms of relative speed or relative density by using Eq. A-71, or

$$V=\rho_0\, u_0\left(\frac{u}{u_0}\right)\left(1-\frac{u}{u_0}\right)=\rho_0\, u_0\left(\frac{\rho}{\rho_0}\right)\left(1-\frac{\rho}{\rho_0}\right) \qquad \text{(A-74)}$$

Differentiating and equating to zero gives the maximum value of volume as

$$V_m=\frac{\rho_0\, u_0}{4}=\rho_m\, u_m \qquad \text{(A-75)}$$

when

$$\rho=\frac{\rho_0}{2}=\rho_m \qquad \text{(A-76}a\text{)}$$

and

$$u=\frac{u_0}{2}=u_m \qquad \text{(A-76}b\text{)}$$

Hence, volume may be expressed as a percentage of its maximum value as

$$\frac{V}{V_m}=4\left(\frac{u}{u_0}\right)\left(1-\frac{u}{u_0}\right)=4\left(\frac{\rho}{\rho_0}\right)\left(1-\frac{\rho}{\rho_0}\right) \qquad \text{(A-77)}$$

and can be referenced to the critical values of ρ_m *and* u_m by

$$\frac{V}{V_m}=\left(\frac{u}{u_m}\right)\left(2-\frac{u}{u_m}\right)=\left(\frac{\rho}{\rho_m}\right)\left(2-\frac{\rho}{\rho_m}\right) \qquad \text{(A-78)}$$

as plotted in Figure A-9, which shows that when relative speed is a point on one arc of the parabola the relative density is a corresponding point on the other arc, since

$$\frac{u}{u_m}+\frac{\rho}{\rho_m}=2 \qquad \text{(A-79)}$$

VEHICLE SPACING

Vehicle spacing may be defined as

$$\delta=1/\rho \text{ mi per vehicle} \qquad \text{(A-80)}$$

The minimum spacing occurs when the speed has been reduced to zero and $\rho=\rho_0$, at which point

$$\delta_0=1/\rho_0 \text{ mi per vehicle} \qquad \text{(A-81)}$$

or the minimum spacing is one vehicle length.

The relative spacing is given by

$$\frac{\delta}{\delta_0}=\frac{1}{\rho/\rho_0}=\frac{1}{1-u/u_0} \qquad \text{(A-82)}$$

Relative speed and relative density are plotted vs relative spacing in Figure A-10. The reference may readily be

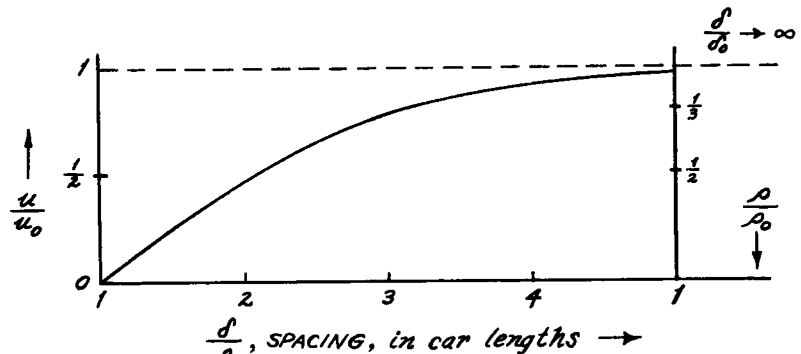


Figure A-10. Relative speed and relative density vs relative spacing.

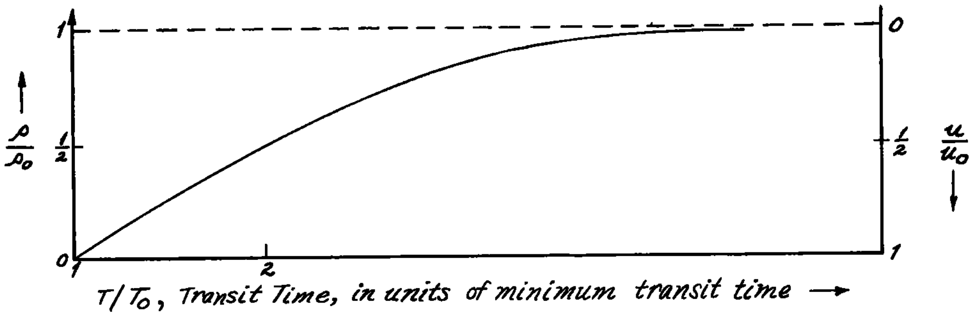


Figure A-11.

changed to the density and speed values corresponding to maximum volume, or

$$\frac{\delta}{\delta_0}=\frac{1}{\rho/2\rho_m}=\frac{1}{1-u/2u_m} \qquad \text{(A-83)}$$

from which it is seen that at maximum volume $u=u_m$ and the spacing is two car-lengths.

TRANSIT TIME

The transit time over a fixed distance, S, is

$$T=S/u \qquad \text{(A-84)}$$

hence,

$$\frac{T}{T_0}=\frac{1}{u/u_0}=\frac{1}{1-\rho/\rho_0}=\frac{1}{1-\delta_0/\delta} \qquad \text{(A-85)}$$

where T_0 is the minimum transit time over distance S, corresponding to maximum speed.

Comparison of Eq. A-85 with Eq. A-82 shows that transit time relates to speed and density in the same way that spacing relates to density and speed. This is plotted in Figure A-11. For convenience the results of Figures A-10 and A-11 are combined in Figure A-12.

TIME HEADWAY

The time headway is defined as the spacing per unit of speed, or

$$\tau=\frac{\delta}{u}=\frac{1}{\rho u} \qquad \text{(A-86)}$$

hence

$$\tau=\frac{60}{V}\ \text{min} \qquad \text{(A-87)}$$

The time headway may also be related to the time spacing at maximum volume, or

$$\frac{\tau}{\tau_m}=\frac{1}{V/V_m} \qquad \text{(A-88)}$$

and the upper abscissa scale of Figure A-9 expresses τ/τ_m as a function of speed and density (note nonlinear scale).

At a speed of $0.4u_m$ the drivers will provide a time headway of approximately 1.5 times that provided at maximum volume; and at a speed of $1.6u_m$ the drivers would provide the same time headway as at 0.4 u_m. If the time headway at u_m is a safe time headway this performance shows a clear inconsistency, inasmuch as a safe headway at u_m

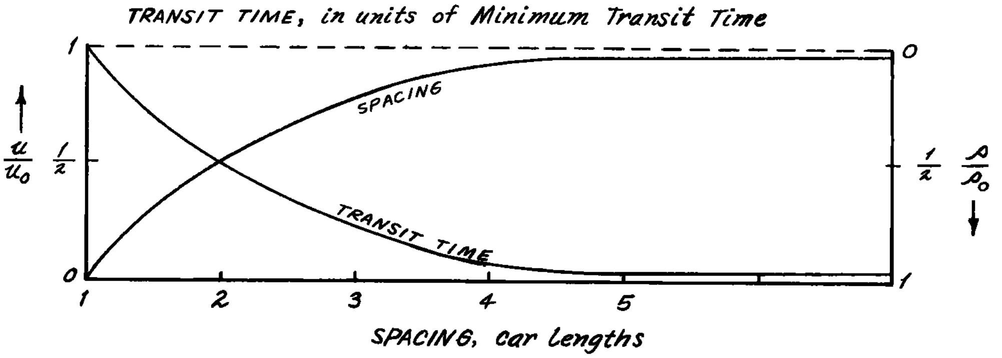


Figure A-12. Transit time and spacing vs relative speed and relative density.

should certainly be a safe headway at $0.4u_m$. On the other hand, if the time headway at $0.4u_m$ and $1.6u_m$ is safe, maximum volume is being attained at unsafe speeds. The reason for this characteristic behavior might be the subject of future investigation.

4. LINEAR THEORY OF TRAFFIC FLOW AND REPLACEMENT BY A VALID TRANSMISSION FUNCTION

THE PROBLEM

Consider a freeway network with entrances $e_1, e_2, \ldots, e_r$ and exits $l_1, l_2, \ldots l_s$. Denote the number of vehicles entering entrance e_1 per unit of time by $e_i(t)$, $i = 1, 2, \ldots, r$; and the number of vehicles leaving exit l_j per unit of time by $l_j(t)$, $j = 1, 2 \ldots, s$. Then the volume of vehicles entering the network, as a function of time, is

$$e(t) = \sum_{i=1}^{r} e_i(t) \tag{A-89}$$

and the volume of vehicles leaving the network, as a function of time, is

$$l(t) = \sum_{j=1}^{s} l_j(t) \tag{A-90}$$

The number of vehicles stored within the network at time t is thus

$$Q(t) = \int_{-\infty}^{t} [e(t) - l(t)]\, dt \tag{A-91}$$

and by direct differentiation it follows that

$$\frac{dQ(t)}{dt} + l(t) = e(t) \tag{A-92}$$

Eq. A-92 is a fundamental equation of macroscopic traffic flow. In its present form, however, it is a statement of an interrelationship that exists between the three variables, Q, l, and e, and cannot be solved. But if a relationship can be shown to exist between any two of these, a differential equation results which is capable of being further analyzed.

It appears reasonable to assume that (macroscopically) the number of vehicles leaving the system per unit of time should be proportional to the number of vehicles within the system, or

$$l(t) = a\, Q(t) \tag{A-93}$$

where a is the constant of proportionality. This is diagrammed in Figure A-13.

Substitution of Eq. A-93 in Eq. A-92 gives

$$\frac{dl(t)}{dt} + a\, Q(t) = e(t) \tag{A-94}$$

which is a linear differential equation that describes the traffic, under the assumption of Eq. A-93. The driving function is $e(t)$. Let it be assumed that

$$e(t) = E\, 1 \tag{A-95}$$

(a Heaviside step function), or that the volume entering the network is constant with time.

Substituting Eq. A-95 in Eq. A-94, taking the Laplace transform, and solving, gives

$$L(p) = \frac{a E}{P(P + a)} = E\left(\frac{1}{P} - \frac{1}{P + a}\right) \tag{A-96}$$

hence

$$l(t) = E(1 - \epsilon^{-at}) \tag{A-97}$$

which is shown in Figure A-14.

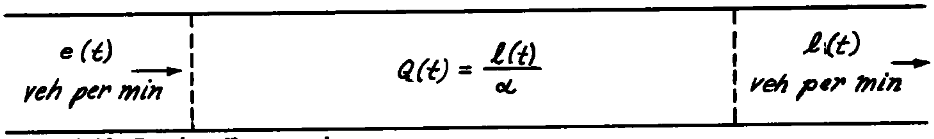


Figure A-13. Road traffic network.

CRITICISM OF LINEAR MODEL

Despite the fact that a model has been "developed," it is not accepted. Consider a section of road with driving function $e_1(t)$ and output $l_1(t)$, with proportionality constant, λ. Assuming $e_1(t) = 1$,

$$l_1(t) = 1 - \epsilon^{-\lambda t} \tag{A-98}$$

Now consider a section of road abutting the line across which $l_1(t)$ is measured. Clearly,

$$e_2(t) = l_1(t) = 1 - \epsilon^{-\lambda t} \tag{A-99}$$

and

$$E_2(P) = \frac{1}{P} - \frac{1}{P+\lambda} \tag{A-100}$$

Hence, from Eq. A-95,

$$\begin{aligned} L_2(P) &= \lambda\left(\frac{1}{P(P+\lambda)} - \frac{1}{(P+\lambda)^2}\right) \\ &= \frac{1}{P} - \frac{1}{P+\lambda} - \frac{\lambda}{(P+\lambda)^2} \end{aligned} \tag{A-101}$$

and

$$l_2(t) = 1 - \epsilon^{-\lambda t} - \lambda t \epsilon^{-\lambda t} \tag{A-102}$$

Again, consider a section of road abutting the line across which $l_2(t)$ is measured. Then

$$e_3(t) = l_2(t) = 1 - \epsilon^{-\lambda t}(1 + \lambda t) \tag{A-103}$$

Hence,

$$L_3(p) = \lambda\left(\frac{1}{P(P+\lambda)} - \frac{1}{(P+\lambda)^2} - \frac{\lambda}{(P+\lambda)^3}\right) \tag{A-104}$$

and

$$l_3(t) = 1 - \epsilon^{-\lambda t}\left(1 + \lambda t + \frac{\lambda^2 t^2}{2!}\right) \tag{A-105}$$

But now assume $e_1(t) = e(t)$ and $l_3(t) = l(t)$. Associated with the entire section of road there is a storage factor, say λ^*. Hence, in response to $e(t) = 1$,

$$l(t) = 1 - \epsilon^{-\lambda^* t} \tag{A-106}$$

However, there exists no λ^* such that Eqs. A-105 and A-106 are equal. Thus, the assumption $l(t) = \alpha Q(t)$, or Eq. A-93, leads to a contradiction. It is concluded that the assumption cannot hold in practice, and should be discarded as a description of macroscopic traffic flow.

The contribution of this analysis is that it has revealed an important requirement.

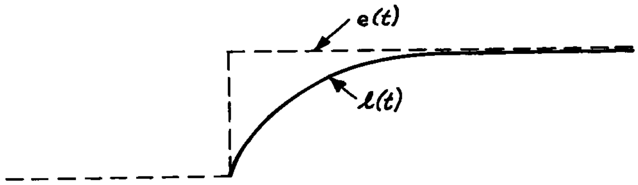


Figure A-14. Entering and leaving functions on assumption $L(t) \propto Q(t)$.

TWO REQUIREMENTS FOR A VALID THEORY

The analysis points to one requirement, here called the "tandem requirement," which can be stated:

If $e_1(t)$ produces $l_1(t) = e_2(t)$, and
if $e_2(t)$ produces $l_2(t)$, then
$e_1(t)$ must produce $l_2(t)$ when considered in the absence of e_2.

The second is the "conservation requirement," that no vehicles are to be created or destroyed within the route network, which is stated mathematically in Eq. A-92.

A model is sought that satisfies both requirements, principles of vehicular flow that no macroscopic theory should violate.

Therefore, the highway now is considered from the viewpoint of a transmission system. A theory is advanced wherein terminal measurements are made of the number of vehicles entering the highway per unit of time and the number of vehicles leaving the highway per unit of time. From these measures an important measure of highway performance is deduced—the transmission function of the highway.

Considered from a probabilistic viewpoint the transmission function is shown to be the probability that a vehicle entering at time τ will leave between $\tau + t$ and $\tau + t + dt$.

TRANSMISSION FUNCTION

Consider a section of highway s miles long with an input of $e(t)$ vehicles per minute and an output of $l(t)$ vehicles per minute, as depicted by Figure A-15. There exists a transmission function, $c(t)$, satisfying the conservation and tandem requirements, such that

$$\int_{-\infty}^{\infty} c(t)\, dt = 1 \tag{A-107}$$

$$E(p)\, C(p) = L(p) \tag{A-108}$$

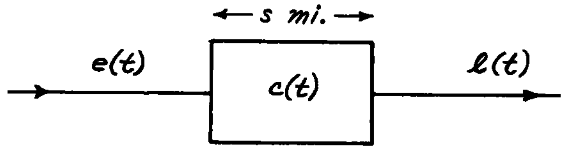


Figure A-15.

where $L(P)$ is the Laplace transform of $l(t) = Ll(t)$, etc. Hence,

$$C(P) = \frac{L(P)}{E(P)} \qquad \text{(A-109)}$$

and

$$c(t) = L^{-1}\frac{L(P)}{C(P)} \qquad \text{(A-110)}$$

Moreover, $l(t)$ is the convolution of $e(t)$ and $c(t)$, or

$$l(t) = \int_{-\infty}^{+\infty} e(x)c(t-x)\,dx \qquad \text{(A-111)}$$

TANDEM REQUIREMENT SATISFIED

Consider two abutting sections of highway—$e_1(t)$, $c_1(t)$, $l_1(t)$ and $e_2(t)$, $c_2(t)$, $l_2(t)$—where $l_1(t) = e_2(t)$, and call $c(t)$ the over-all transmission. From Eq. A-109,

$$C_1(P) = \frac{L_1(P)}{E_1(P)} \qquad \text{(A-112)}$$

and

$$C_2(P) = \frac{L_2(P)}{E_2(P)} \qquad \text{(A-113)}$$

but then

$$C(P) = C_1(P)\ C_2(P) \qquad \text{(A-114)}$$

Hence, the over-all transmission is the convolution of the individual transmissions, or

$$c(t) = \int_{-\infty}^{+\infty} c_1(x)\ c_2(t-x)\,dx \qquad \text{(A-115)}$$

This is readily extensible, so

$$C(P) = C_1(P)\ C_2(P)\ \ldots\ C_k(P) \qquad \text{(A-116)}$$

and thus completely satisfies the tandem requirement.

DISCUSSION OF TRANSMISSION FUNCTION

The transform of $c(t)$ is, in theory, representable by a physical electrical network. It may then be interpreted as:

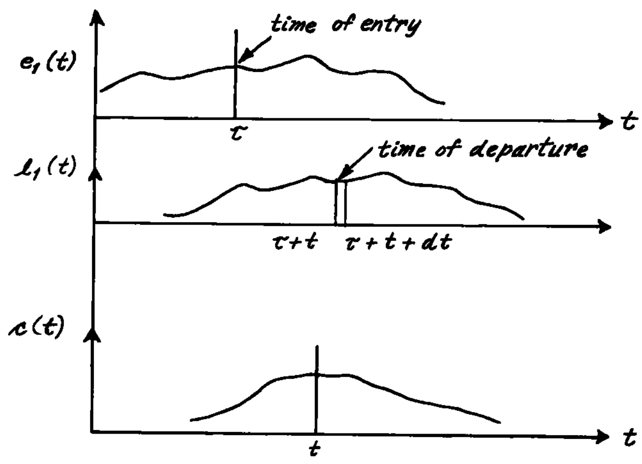


Figure A-16. Input, output, and transfer functions.

"over the time interval of observation the highway performed as if it were this network." Over a different time interval, and under different conditions, the network would be expected to change. It might then be possible to correlate factors such as weather, time of day, state of disrepair, communications to drivers, etc., with changes in network parameters.

One interpretation of $c(t)$ that is of overriding interest and importance is well understood by writers in other fields (such as Brownian motion). That is, $c(t)$ is the probability that a vehicle that enters the highway at time τ will leave between $\tau + t$ and $\tau + t + dt$ (Fig. A-16). Hence, the probability that a vehicle will leave the system within T seconds after entering is

$$P\{t \leq T\} = \int_{-\infty}^{t} c(t)\,dt \qquad \text{(A-117)}$$

It is clear that an important function has been deduced that will lend insight into the mechanism of highways. It is based on elementary measurements, readily obtainable. It is extensible to complicated highway networks, since $e = \Sigma\, e_i$ and $l = \Sigma\, l_i$. From a study of the nature of $C(p)$ and $c(t)$, predictive doctrines may be deducible.

If the highway is to be studied practically to the end that the most meaningful results will be derived, another important property is homogeneity. Homogeneity is assured in light of the theory if the distribution of vehicular velocities is independent of the point of measurement. This might be a subject for further investigation.

5. THE TRANSMISSION FUNCTION AS A MODEL FOR STUDY OF TRAFFIC NETWORKS

An extensive literature search reveals that descriptive theories concerning vehicular traffic are inadequate or are restricted to very limited situations. In general, the following classifications of theories are accepted:

1. Analytic and deterministic model—wherein the characteristics of the vehicles are considered under assumed driver behavior.
2. Queue theory treatment of a stochastic model—which

necessitates limitations of points of entry into the system.

3. Traffic flow in a continuum—based on an assumed analog to molecules in semicompressible fluid flow.

4. Monte Carlo integration or simulation model—preceded by a thorough analysis of observed traffic data in order to understand and simulate the real traffic complex.

PRESENT APPROACH

The objection to all of the foregoing theories is that they appear analogous to trying to explain the performance of a complex piece of electronic gear without first having determined the performance characteristics of a resistor, an inductor, or a capacitor.

In category 1 the driver behavior is assumed; in category 2 one assumes limited points of entry; in category 3 one assumes an analog; and in category 4 one assumes characteristics to enable simulation. For situations in which these assumptions do not hold, it is assumed that the model does not hold.

In the present approach nothing is assumed. However, it is stipulated that the two requirements from the previous section must be satisfied; that is,

1. Conservation of vehicles—no "sources" or "sinks" in the traffic complex.
2. Tandem requirement—the theory that applies to an elementary section of roadway, properly employed, must apply to two or more sections in tandem.

Thus, to the previous list of four models is added a fifth, as follows:

5. Transmission function model—which considers a traffic network as a natural phenomenon to be investigated by terminal performance measurements.

It is suggested that simple networks be studied first and the variations of weather, peripheral parked cars (marginal friction), road surface, and other factors (such as communications to drivers) later be included as variations in the network transmission function. It is expected that:

1. From the measurements-validated performance of elementary networks, a knowledge will ensue about more complex systems.
2. From variations among the same system characteristics, an indication of normal or abnormal behavior will become apparent.
3. From similarities between different systems, a general theory will emerge.
4. From all the foregoing, predictive theories founded on practice will emerge.

TRAFFIC NETWORK MODEL

A traffic network is comprised of a bounded section of roads (the boundary corresponding to a simple closed curve drawn on a map) where accounting is made for all vehicles crossing the boundary. For example, Figure A-17 depicts two types of traffic network and their boundaries.

ENTERING AND LEAVING FUNCTIONS

In the previous section it is shown that the movements of vehicles in and out of a network can be expressed by the entering function

$$e(t) = \sum_{i=1}^{r} e_i(t) \qquad \text{(A-89)}$$

and the leaving function

$$l(t) = \sum_{j=1}^{v} l_j(t) \qquad \text{(A-90)}$$

These functions are obviously suitable descriptors for any network. In addition to their applicability to the more complex networks of Figure A-17, their use is readily appreciated in regard to the more elementary networks of Figure A-18, where in (a), the most basic network, for a simple section of road,

$$e(t) = e_1(t) \text{ and } l(t) = l_1(t) \qquad \text{(A-118)}$$

and in (b), for the merge of two freeways,

$$e(t) = e_1(t) + e_2(t) \text{ and } l(t) = l_1(t) \qquad \text{(A-119)}$$

It is clear that similar formulations hold for more complex networks and, indeed, for all networks.

Also, Eqs. A-91 and A-92 express the conservation requirement. The differential form is repeated here

$$\frac{dQ(t)}{dt} + l(t) = e(t) \qquad \text{(A-92)}$$

The tandem requirement can be expressed as follows: In the first section, if $e_1(t)$ results in $l_1(t)$ and, in the second section, if $e_2(t)$ results in $l_2(t)$ and, in general, in the nth section, if $e_n(t)$ results in $l_n(t)$; and further, if $l_j(t) = e_{j+1}(t)$, for $j \geq 1$; then $e_1(t)$ must result in $l_s(t)$, if there are s sections.

DISCUSSION OF FUNDAMENTAL EQUATION

Eq. A-92 is a differential relationship involving the three variables, $Q(t)$, $l(t)$, and $e(t)$. As it exists, of course, it cannot be solved. If one of the variables can be shown to be functionally related to one of the others, the resulting differential equation is amenable to analysis and solution. For example, Figure A-19 depicts the resulting storage on the assumption that $e(t)$ is a Heaviside unit step function, corresponding to a constant volume of vehicles entering after time $t = 0$, and $l(t)$ has various values as indicated. (In practice, of course, $e(t)$ and $l(t)$ are to be measured, not assumed. It is only through measurement that the physics of the basic interrelationships is learned.)

In a more general sense it may be said that the cause is the entering volume, $e(t)$, and the desired effect is the leaving volume, $l(t)$, while the undesired effect (on freeway networks and their environs, to which this study is restricted, but generally on any network whose purpose is transportation and not parking) is stored vehicles, as schematically represented in Figure A-20.

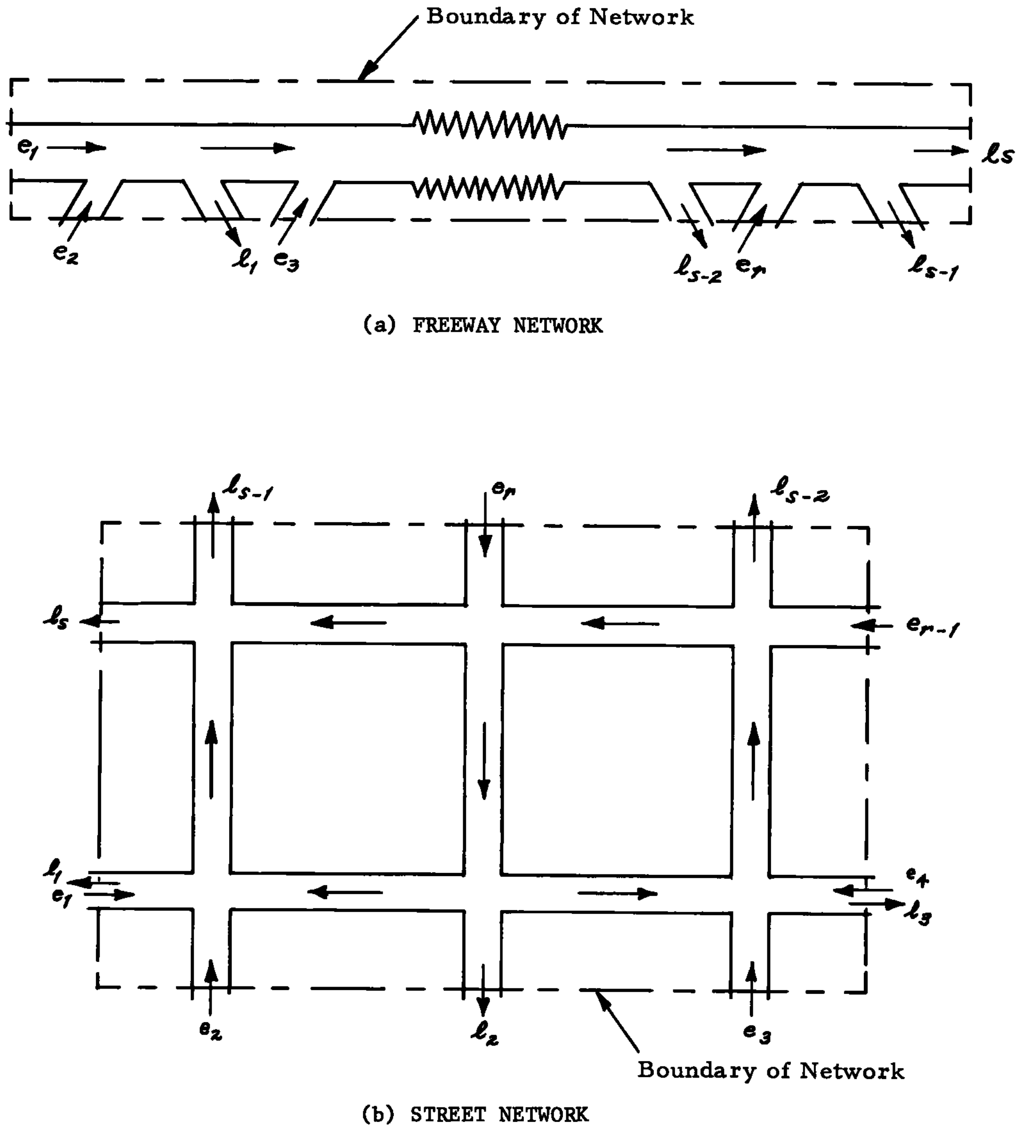


Figure A-17. Examples of traffic networks.

It is the purpose of network refinements, such as roadway and traffic control improvement, to have $l(t)$ approach $e(t)$ as closely as possible and to try to eliminate $Q(t)$. In a sense, then, travel time is lost time representing the resistance of a network to one's desire to get from here to there, and only due to the physical requirement that one pays for travel with time is there a vehicle storage in the system. (The minimum cost of 130 miles on the New York Thruway is $130/65 = 2$ hr, and each user stores his vehicle on the 130-mile stretch for a minimum of 2 hr. The cost rises, as well as the storage, when the section carries heavy traffic).

Continuing the discussion of Eq. A-92, another functional relationship, one between Q and l, was tried in the previous section and found to be inadmissible. It was that

$$l(t) = a\, Q(t) \qquad \text{(A-93)}$$

When tested for two abutting portions of a highway, with $e(t)$ as a step function, the solutions for $l(t)$ were contradictory. Failure of this form of functional relationship led to the notion of the transmission function model.

TRANSMISSION FUNCTION CONCEPT

The expression of the transmission was developed in the previous section. The derivation is documented here more fully in order to provide continuity for the follow-up discussion.

Attention is directed to an elementary section of highway as the building block for this theory. Let the number of vehicles entering a length of highway of S miles be $e(t)$ vehicles per minute and the number of vehicles leaving be $l(t)$ vehicles per minute, as shown in Figure A-21.

The highway may be considered to be a unit that converts, or transforms, time functions from one form to another. In Figure A-22 $e(t)$, as input, produces $l(t)$ as output; hence, $T[e(t)] = l(t)$, or $l(t)$ is a transformation of $e(t)$.

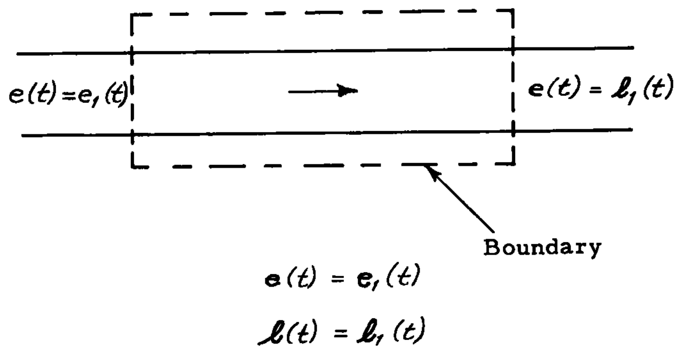
Further, the transformation is brought about due to the interaction of vehicles with other vehicles and with the properties of the roadway. These properties vary with such parameters as time of day, weather, road surface, and visibility, and as such are difficult to describe as a group. However, the properties of the roadway cannot violate the fundamental requirements previously summarized. These requirements lead to a mathematical description in the context of which all other variants can be studied, as will become apparent.

In what follows, $F(p)$ is used to designate the Laplace transform of $f(t)$, employing the customary device of pairing upper- and lower-case letters.

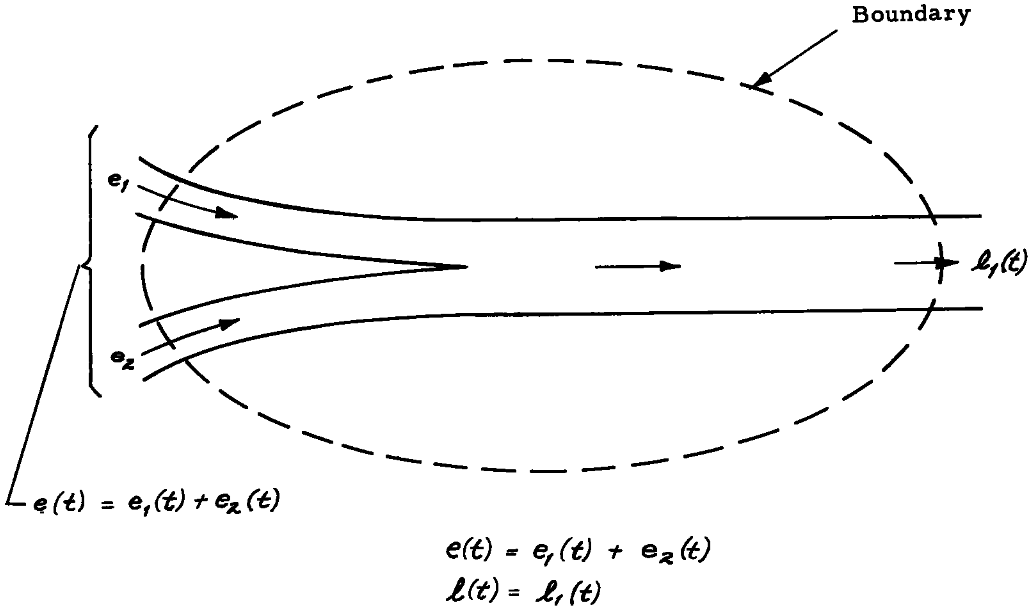
For this discussion, a transmission function, $c(t)$, is defined such that

1. $c(t) = 0$ for $t < 0$ (A-120)
2. $$\int_0^\infty c(t)\, dt = 1 \qquad \text{(A-121)}$$
3. $$C(p) = \frac{L(p)}{E(p)} \qquad \text{(A-122)}$$

As pointed out earlier, the second of these conditions satisfies the requirement for the conservation of vehicles, and the third satisfies the tandem requirement.



(a) SIMPLE SECTION OF ROAD



(b) MERGE OF TWO FREEWAYS

Figure A-18. Application of e(t) *and* l(t) *to more basic networks.*

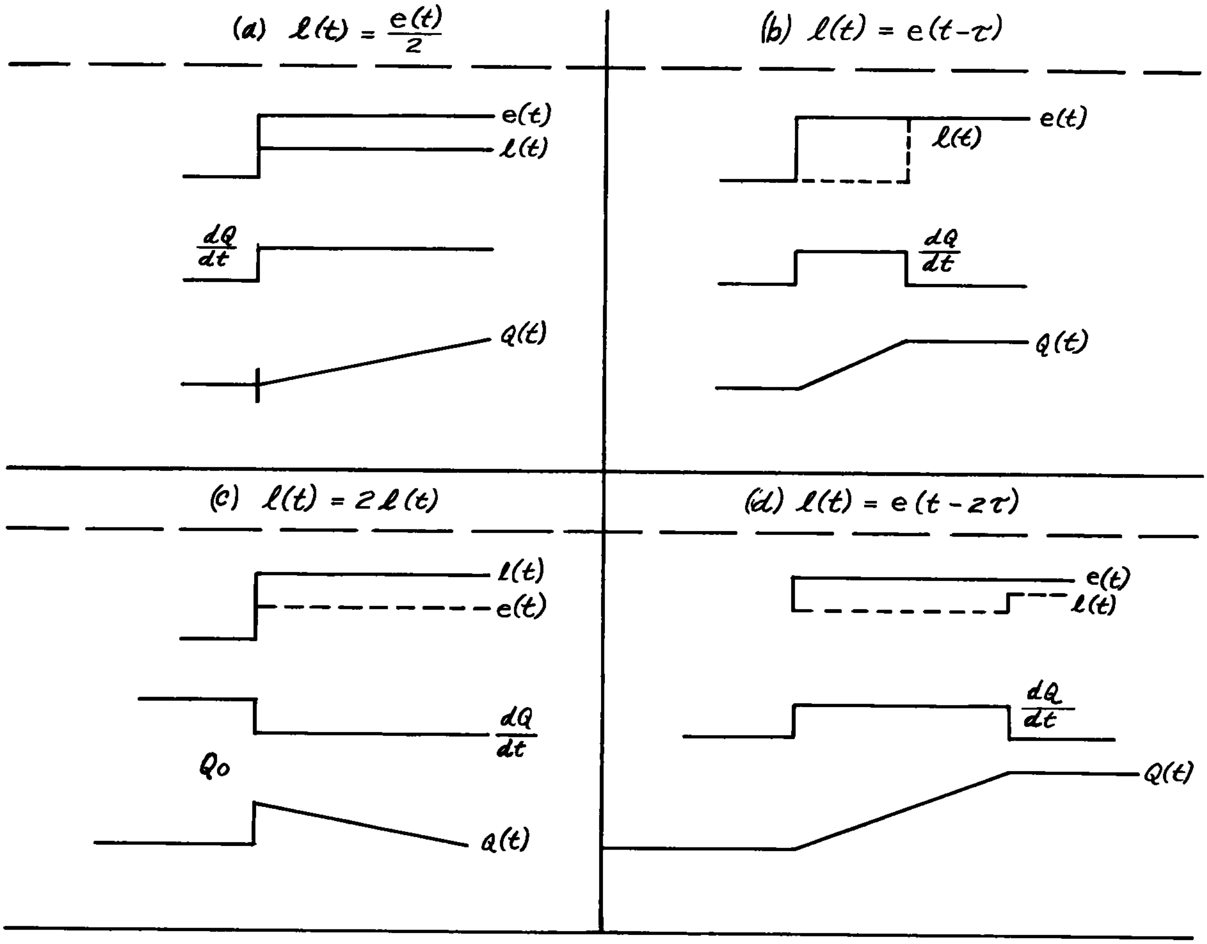


Figure A-19. Examples of law of conservation of vehicles under indicated assumptions.

NETWORK INTERPRETATION OF TRANSMISSION FUNCTION CONCEPT

The Laplace transform of the transmission function can be represented by a physical network, the electrical analogue of which is now considered. To illustrate this concept two situations are treated, neither of which presumes to represent the physical situation, but each of which provides insights. (The true situation will only be revealed through measurement.)

Consider the elementary section of highway shown in Figure A-21 and its performance, as follows:

1. Ideal performance—in which the only storage that takes place is due to the nominal trip time.

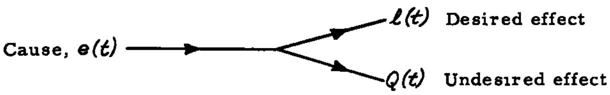


Figure A-20. Network depicted as cause and effect.

2. Probable performance—in which vehicular interaction causes further delay.

In each instance $c(t)$ is shown as a time function and the electrical network analogue is as implied by $C(p)$.

Ideal Performance

Ideal performance would involve a travel time delay of τ minutes, so

$$l(t) = \begin{Bmatrix} e(t-\tau), \text{ for } t > \tau \\ 0, \text{ for } t \leqq \tau \end{Bmatrix} \tag{A-123}$$

hence,

$$L(p) = \int_0^\infty e^{-pt}\, e(t-\tau)\, dt \tag{A-124}$$

Let $t = \tau + \theta$, and let θ be the variable of integration in Eq. 124, or

$$L(p) = e^{-p\tau} \int_0^\infty e^{-p\theta}\, e(\theta)\, d\theta \tag{A-125}$$

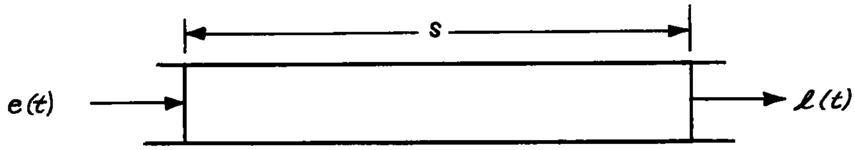


Figure A-21. Elementary section of highway.

Further,

$$E(p) = \int_0^\infty e^{-pt}\, e(t)\, dt \tag{A-126}$$

$$C(p) = \frac{L(p)}{E(p)} = e^{-p\tau} \tag{A-127}$$

and

$$c(t) = \delta(t - \tau) \tag{A-128}$$

The last is a Dirac function.

This result is shown in Figure A-22, where it is observed that the output, $l(t)$, is an exact replication of the input, $e(t)$, following the travel time delay of τ minutes.

The ideal performance may be represented by an electrical network. A delay of τ minutes is introduced by a delay line and output (current) follows input (current) because the output load is representable by a pure resistance of 1 ohm as shown in Figure A-23. It is seen that a one-to-one correspondence exists and that $e(t)$, the entrance function, corresponds to the input current into the network and $l(t)$, the output function, corresponds to the output current into a 1-ohm load.

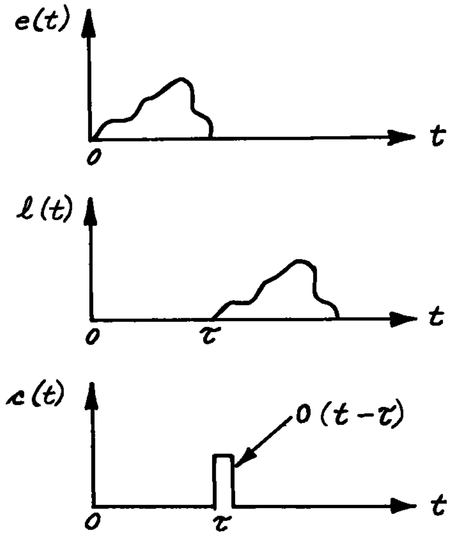


Figure A-22. Ideal performance of traffic network.

Probable Performance

In practice it is clear that the pure performance of the Dirac function will not take place. Rather, one might expect "shunt capacity across the load" to intrude into the situation.

Assume the *form* of the transmission is:

$$g(t) = \begin{Bmatrix} Ke^{-a(t-\tau)}, \text{ for } t > 0 \\ 0, \text{ for } t \lesssim 0 \end{Bmatrix} \tag{A-129}$$

It is required that

$$\int_\tau^\infty c(t)dt = 1 \tag{A-130}$$

Hence, for $g(t)$ to be representative of a transmission function, since

$$K\int_\tau^\infty e^{-a(t-\tau)}\, dt = \left[Ke^{a\tau}\frac{e^{-at}}{-a}\right] = \frac{K}{a} = 1 \tag{A-131}$$

it is required that $K = a$. Therefore,

$$c(t) = \begin{Bmatrix} ae^{-a(t-\tau)}, \text{ for } t > 0 \\ 0, \text{ for } t \lesssim 0 \end{Bmatrix} \tag{A-132}$$

represents a possible transmission function. Now,

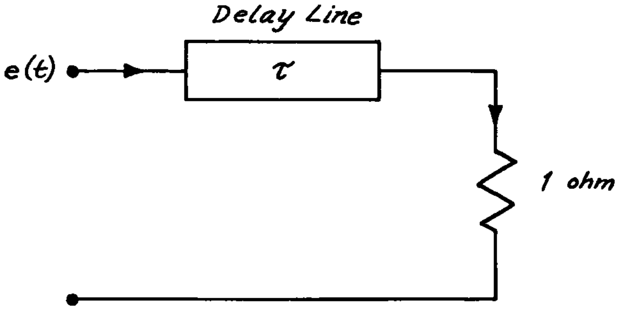


Figure A-23. Electrical network analogue of ideal performance.

$$\begin{aligned} C(p) &= ae^{a\tau}\int_\tau^\infty e^{-(p+a)t}\, dt \\ &= ae^{a\tau}\frac{e^{-(p+a)\tau}}{p+a} \\ &= \frac{ae^{-p\tau}}{p+a} \end{aligned} \tag{A-133}$$

Also, assuming $e(t)$ = unit step = 1,

$$E(p) = \frac{1}{p} \tag{A-134}$$

and

$$L(p) = E(p)\,C(p)$$
$$= \frac{a\,e^{-pT}}{p(p+a)} = \frac{e^{-pT}}{p} - \frac{e^{-pT}}{p+a} \tag{A-135}$$

Then, clearly,

$$l(t) = \begin{cases} 1 - e^{-a(t-\tau)}, & \text{for } t > \tau \\ 0, & \text{for } t \leq \tau \end{cases} \tag{A-136}$$

Eqs. A-132, A-134, and A-135 are depicted in Figure A-24.

The network analogue is shown in Figure A-25. It is seen that the output is again the current into the load of 1 ohm. However, the additional storage required by slowdown due to traffic (delaying the fulfillment of purpose in getting current into the load) is represented by a shunt capacitance across the load of $1/a$ farads. (In practice, of course, one must measure $e(t)$ and $l(t)$, determine $c(t)$, and synthesize the network based on $C(p)$.)

PROBABILITY INTERPRETATION OF TRANSMISSION FUNCTION

It has been shown that

$$\int_0^\infty c(t)\,dt = 1 \tag{A-121}$$

If $c(t)$ is nonnegative, it may be interpreted as a probability density function with the following important property:

The probability, P, that a vehicle will leave the network at a time less than or equal to T is

$$P(t \leq T) = \int_{-\infty}^{T} c(t)\,dt \tag{A-137}$$

However, due to time constraints, it is not known whether or not the transmission function is nonnegative, even though it is known that this property is shared by both the entering and the leaving functions. This probably represents a fruitful area for future investigation.

PRACTICAL APPLICATION

Having developed a theory the question arises as to its applicability. Specifically, having obtained $e(t)$, how does it respond to the following criteria:

1. Is it objective?
2. Is it practical?
3. Is it a valid measure?
4. Is it reliable?
5. Is it feasible?

Objectivity

The transmission function, as has been shown, depends for its calculation on continuous measurements of $e(t)$ and $l(t)$, the entering and leaving volumes of a traffic network. *Per se*, then, it is an objective measure.

Nevertheless, a particular aspect of it will be subjective. Assuming a quiescent system at $t = 0$, successive integration by parts evaluated at time T reveals

$$C(p) = \frac{e(T) + \frac{e'(T)}{p} + \frac{e''(T)}{p^2} + \ldots}{l(T) + \frac{l'(T)}{p} + \frac{l''(T)}{p^2} + \ldots} \tag{A-138}$$

Thus, the degree of derivative variation over the most elementary time interval will determine the complexity of the electrical network. This is dependent on the time interval over which "instantaneous" volume measurements are averaged. The question must be empirically resolved, and experience will provide guidelines on which standards of measurement can be based.

Validity

In the philosophical sense that the purpose of the roadway is to transmit drivers from here to there, the transmission function is a valid measure of performance of roadway function. In the public sense that the measure must account for all drivers in the system, the transmission function integrates the individual performances and ascribes an over-all function, and so is a valid measure. In the etiological sense that one must relate phenomenological cause and effect, the transmission function measures exactly these and describes a function that recreates effect in the presence of cause, and so is a valid measure. In the scientific sense that an elementary theory should contain within itself the building block of the complex, the transmission function accomplishes precisely this, and so is a valid measure. In the practical sense that the measurements envisioned are well within the state of the art, the transmission function is a valid measure. In the ecological sense that the transmission function provides a measure of mutual relationship of the human and his environment, it is a valid measure.

Practicality

The practicality of the transmission function has been touched upon in terms of the state of the art. But what is the potential usefulness of the transmission function as a research tool? By correlating conditions on the same road with differing transmission functions one has a relative measure of difference. For example, how does inclement weather affect the transmission function of a roadway? How is the transmission function affected by shutting down a single lane of a three-lane highway? How is the transmission function of roadway A affected by initiating service on (new) roadway B? How is the transmission function affected by communications to drivers? How is the transmission function affected by a change of speed limit on a freeway? By the proximity of parked cars?

In time one moves to esoteric interpretations based on transmission function experience such as: that speeds of so many miles per hour will enhance the performance of

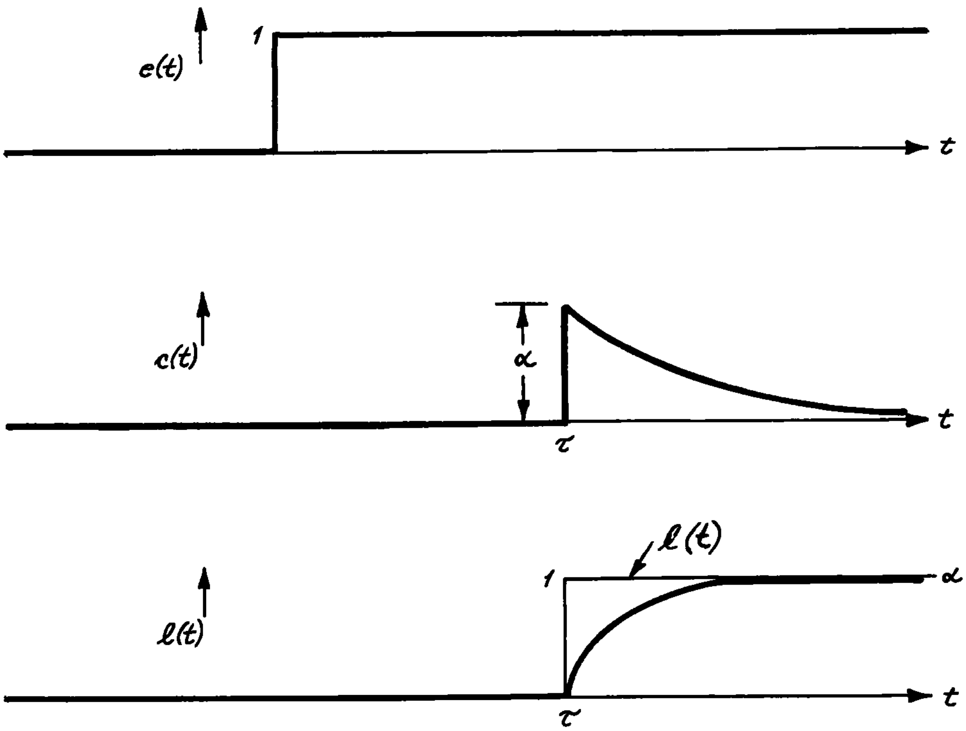


Figure A-24.

a roadway; that cars should be parked so many feet from a roadway; that brilliant roadside lights of particular sorts blind drivers and slow traffic.

In short, it is a tool that can be used for studying and improving. It is practical.

Reliability

It is clear that the reliability requirement must be satisfied to enable the interpretation of roadway phenomena in terms of transmission function data.

The reliability of the transmission function as a measure of traffic behavior must be borne out by practice. Some questions that arise are:

1. Is it a stable duplicable measure: is the roadway, under roughly identical conditions, portrayed by a roughly identical transmission function?
2. Is it a sensitive measure: do important variations in behavior pattern manifest themselves adequately to be observed in the transmission function? What constitutes a proper stretch of roadway so that it does become so?
3. What time interval of integration is required (see prior discussion under "Objectivity") so that items 1 and 2 above are both optimally satisfied?

Feasibility

The following outlines a set of elementary experiments that might be performed to provide an answer to the question: Is it feasible?

1. Choose a relatively simple section of roadway to permit understanding the nature of the element as the key to the complex.
2. Record the roadway geometrics, log weather, road surface, unusual facts about road repair, encumbrances, etc.
3. Measure $e(t)$ and $l(t)$ continuously.
4. Compute $C(p)$ and $c(t)$, as previously described. (Hence, the transmission function is recorded as a function of time and as a network with time-varying parameters.)

Attention might now be directed to the following questions:

1. What is the day-to-day correlation of $c(t)$ with itself; other things roughly being equal, is $c(t) \cong c(t-\tau)$?
2. How does the averaging interval (see "Objectivity") affect the result?

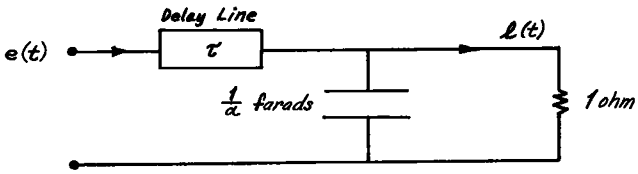


Figure A-25. Electrical network analogue of probable performance.

3. How does rain (wet roads, lowered visibility) affect $c(t)$; i.e., compare $[c(t)]_{rain}$ with $[c(t)]_{no\ rain}$.

4. How does $c(t)$ respond to variations in network parameters? Can the sensitivity be improved by changing the averaging interval? (see "Objectivity" and "Reliability").

Indeed, one may now study:

5. How does $c(t)$ vary on the same road with and without communications to drivers?

Having carried the work to a point where the performance of the transmission function is reliable and sensitive it can now be compared in different roads in:

6. How does $c(t)$ vary from road to road with the same geometrics; i.e., compare $[c(t)]_{road\ 1}$ with $[c(t)]_{road\ 2}$, with other parameters as nearly equal as possible.

APPENDIX B

FREEWAY SURVEILLANCE STUDIES

1. APPLICATIONS OF REAL HIGHWAY DATA TO TRAFFIC CONGESTION COMPUTATIONAL LOGIC

FIRST APPLICATION

The ultimate objective of this project was to determine whether or not traffic operations can be improved by the use of surveillance methods and communications with drivers. Because there are no generally agreed upon criteria for the evaluation of traffic operations, it is difficult to determine when they are being improved except, perhaps, by intuition. Nevertheless, it has been suggested that a possible means of improving traffic operations is to reduce congestion or traffic jams by predicting their occurrence and advising drivers to avoid these areas. Improvement of traffic operations by this method requires a system of several components—namely, a definition of traffic jam or congestion; a means of communication with drivers; a surveillance system capable of providing accurate up-to-the-minute data on traffic conditions; and a model to predict traffic congestion.

This research concerns one of the components—the development and evaluation of a model to predict traffic congestion. Inasmuch as the model will be used in a real-time surveillance and control system, it must meet some basic requirements in addition to predicting congestion accurately. The model must be capable of predicting traffic congestion far enough into the future to allow controls to be executed; it must be possible to calculate the predictions within about one minute; and the model must use only data that are readily available either from the surveillance system or from historical data.

The model was developed from experience, intuition, analysis of data and results obtained from a surveillance system installed on the Congress Street Expressway in Chicago, and information contained in traffic reports and articles.

The goal of any research should be twofold. The first part is specific and obvious: to satisfy the immediate objective of the program (in this case, to determine whether surveillance and communications can be used to improve traffic operations). The second part is general and perhaps less obvious: to develop new methods and concepts of solutions to unsolved problems or to extend the range of useful application of existing methods for solutions.

Source of Data

The Congress Street Expressway pilot detection system consists of traffic detectors on the ramps and at selected locations along the westbound section of the expressway between Cicero Avenue in Chicago and First Avenue in Maywood. The detectors are either sonic presence detectors or sonic motion detectors. Volume, speed and occupancy were measured every minute at various locations and the data were recorded on punched paper tape. The model described here was developed from the data concerning the East Avenue to Harlem Avenue section, which is approximately 0.85 miles in length.

Initial studies of the Congress Street Expressway data yielded some relationships between volume, speed, and density that are of importance in this discussion. Analysis of the data has shown that for densities above 30 vehicles per mile per lane a linear relationship exists between speed and density. The main observation is that above about 50 vehicles per mile per lane, the volume and speed decrease with an increase in density. As density increases this trend continues until both volume and speed reach zero. On the basis of these relationships, congestion is defined to be a density greater than or equal to 80 vehicles per mile per lane, which should have forced speed to reduce to 18.7 mph.

The distribution of traffic volume by lane is somewhat dependent on the total hourly volume, but also has considerable fluctuations in it. Only the trend of the distribution with hourly volume can be described. During light traffic of less than 1,000 vph, lane 1 (median lane) carries less traffic than either lane 2 or lane 3. For hourly volumes of 3,000 to 4,000 vph, lane 1 and lane 2 each carry about 40% of the traffic. At high volumes of 5,000 vph or above, lane 1 and lane 2 each carry about 35 to 40% of the traffic.

Technical Approach

The average density on a section of freeway at a given time is a function of the input volume and the output volume during the previous time interval and the density at the beginning of the previous time interval. Thus, the density in vehicles/mile/lane for the section under discussion here is

$$R_t = R_{t-1} + \frac{[(V_i)_t - (V_o)_t - (R\,V_o)_t]}{n\,L} \qquad \text{(B-1)}$$

in which

R_t = density at time t;
R_{t-1} = density at time $t-1$:
$(V_i)_t$ = number of vehicles into the study section during the interval $t-1$ to t;
$(V_o)_t$ = number of vehicles out of the study section via the main expressway during the time interval $t-1$ to t;
$(RV_o)_t$ = number of vehicles out of the study section via the Harlem off-ramp during the time interval $t-1$ to t;
n = number of lanes; and
L = length of the expressway study section, in miles.

Therefore, to predict the density at a time greater than t by this formula it is necessary to predict V_i, V_o and RV_o. It is also necessary to know an initial value of the density and V_o.

The following data for the study section are available from the surveillance system for each minute:

1. The volume count in each of the three lanes at East Avenue.
2. The volume count on the off-ramp at Harlem Avenue.
3. The volume count on the on-ramp at Harlem Avenue.
4. The volume count for two lanes only at Harlem Avenue.
5. The volume count in each of the three lanes at DesPlaines Avenue.
6. The speed in each of the lanes at DesPlaines Avenue.

The initial density for the study section was determined by taking the average density at DesPlaines Avenue and projecting it back in time to the study section, DesPlaines being the only location where speed data are available for all three lanes. The initial density for this section is given by

$$(R_S)_t = (R_{DP})_{t+1} - \frac{(RV_i)_{t+1}}{3 \times 0.55} + \frac{(RV_o)_t}{3 \times 0.85} \qquad \text{(B-2)}$$

where

$(R_S)_t$ = density in study section at time t;
$(R_{DP})_{t+1}$ = average lane density at DesPlaines Ave. at time $t+1$;
$(RV_i)_{t+1}$ = number of vehicles into section by Harlem on-ramp during time t to $t+1$;
$(RV_o)t$ = number of vehicles out of section by Harlem off-ramp during time $t-1$ to t; and

density is computed by dividing volume by speed.

This method has been used to determine the initial density for lack of better data. During a real-time surveillance study, the initial density should be determined by a manual count or estimated by the foregoing method at a time when the volume is very low. These densities were calculated for every minute. They seem to give reasonable values, which vary from a low of about 20 veh/mile/lane at 3:00 PM to a momentary high of 112 veh/mile/lane at about 6:00 PM.

Data on volume counts at Harlem were available for two lanes only. One volume was about 30 vpm and the second about 20 vpm, identified as v_1 and v_2, respectively. Lacking a complete count at Harlem, several attempts were made to estimate V_o accurately enough to satisfy the needs of the computational logic. These are described in the following.

Application of the Technique to the Data

The first attempt to estimate V_o was to set

$$(V_o) = r_1\,(v_1)_t \qquad \text{(B-3)}$$

where $r_1 = 2.5$. This assumes that v_1 is 40% of the total count. The densities calculated for this value of V_o resulted in negative densities after about 10 min. Substituting

$$(r_1\,v_1)_t = (V_o)_t \qquad \text{(B-4)}$$

in Eq. B-1, summing over $t=0$ to $t=T$, and rearranging terms, gives

$$R_T - R_o = \frac{1}{n\,L}\sum_t\,(V_i - RV_o) - \frac{r_1}{n\,L}\sum_t\,(v_1) \qquad \text{(B-5)}$$

all the terms of which are known except r_1 and R_T. Thus, using r_1 as the variable parameter R_T can be forced to be any number desired. The final density obtained by extrapolating back in time was about 20 veh/mile/lane. Setting R_T equal to this value and solving for r_1 gives $r_1 = 2.438$. Densities calculated using this value are negative for some time intervals. From the results of the four hours of data, the value of r_1 that will insure a minimum density of zero will give a maximum density of 221 veh/mile/lane. Hence, there is no satisfactory way for estimating V_o from v_1 alone by using a constant multiplier. This is due to the variability in lane volume distribution, as shown in Figure B-1.

The second attempt to estimate V_o was to set

$$(V_o)_t = r_1\,(v_1)_t + r_2\,(v_2)_t \qquad \text{(B-6)}$$

Using r_1 and r_2 as variable parameters, the value of R, the density, can be set equal to any desired value at two points in time. This was done with the extrapolated values at about 3:45 PM and 7:00 PM. Solving the resulting simultaneous equations gave $r_1 = 1.11370$ and $r_2 = 1.96442$. The densities calculated from these values of r_1 and r_2 are fairly reasonable, varying from a low of about 7 veh/mile/lane at about 3:00 PM to a high of about 130 veh/mile/lane at about 5:30 PM. The calculated values are close to

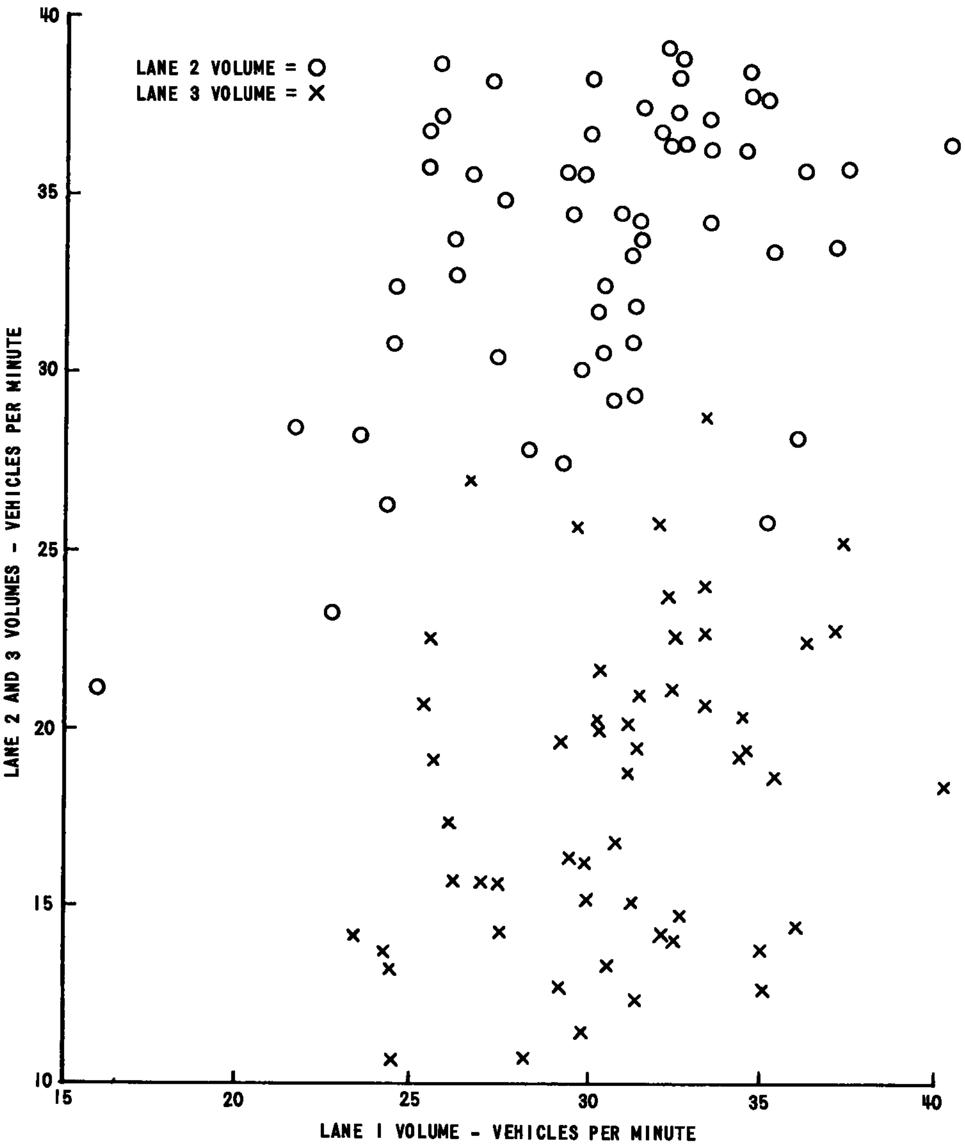


Figure B-1. Distribution of lane volumes on Congress Street Expressway, Chicago, at East Avenue, 3:00 to 4:00 p.m.

$r_1 = 1$ and $r_2 = 2$; however, any change in these values produces large changes in the density values.

Although it is recognized that the values of the coefficients r_1 and r_2 obtained by curve fitting to the present data do not make them satisfactory for application to all data taken at Harlem, justification for their use here is based on the fact that in an operational system data would be taken on all lanes, which would provide a better accuracy in measuring volume. Thus, for the purpose of maintaining a running tally on the density in the test segment, V_t is read from the detectors at East Avenue, ramp volume "out" at Harlem is read from that detector, and V_o of Harlem mainline is computed from the readings of two detectors by the use of the above equation.

The problems associated with acquiring and processing data for prediction of volumes are now discussed. The values of the variables V_t, V_o, and RV_o may be predicted on the basis of historical data or on present (and/or recent) data measurements or on a combination of the two types. The use of historical data alone will give the values as a function of chronological time only. Application of present measurements to the prediction of V_0 makes use of the relationship between volume and density.

For instance, the average value of each of the variables as a function of time may be determined from several sets of data for typical days. This value could then be used as a prediction of the variable at time t. Although the distribution of these values is consistent (that is, the shape of the curves for different days is the same) over the sets of data, the actual values may vary to a significant degree.

This causes errors in the predictions. Some of this error may be removed if a combination of historical and present data is used. The prediction could be based on the present value of the variable and the trend of the values as indicated by the historical curve for time t. Both methods require a historical table or curve for each section of highway on which predictions are to be made.

Otherwise, the predictions might be based on current measurements of the volumes and densities and the known relationships between them. It is expected that this method could be more easily applied to other highways without major modifications to make it suitable to the locale. The accuracy and ease with which these variables can be predicted is dependent on the amount of advance information known about the variable. In this particular case values of V_i will be more difficult to predict than V_o or RV_o because of the sparse instrumentation upstream of East Avenue, which limited volume counts to one lane only at each of the stations where data were taken.

The experience previously noted in attempting to compute V_o at Harlem based on a fixed distribution for the three lanes and data on only one lane indicated the very small likelihood of success in predicting V_i from volumes taken upstream of East Avenue, particularly where the expressway widens to four lanes. A preliminary plot of the total volume at East Avenue and one lane each at Cicero and Austin showed that there was no strongly correlated appearance. Therefore, values of V_i are based on the slope of the volume curve at East Avenue, historically derived, and the current value of V_i. The volume curve used is averaged from the hourly volume curves for a period of 21 days at East Avenue.

The exit ramp volume at Harlem, RV_o, also could be based on historical data. However, no definable pattern of flow was apparent by visual examination of the data; certainly nothing showing any kinship to the volumes passing along the main stream of the expressway. Actually it appeared that the distribution of volumes would be well represented by an average value without having any appreciable effect on the values of V_o. This would offer a simple solution to the representation of RV_o in an automated warning system. But neither of these methods has any appeal, from a philosophical viewpoint. Intuitively, it would be expected that the ramp exit volume should have some dependency on the volume in the main stream, particularly in view of the fact that the main stream is sensed according to individual lanes, and it might be expected that the traffic under the East Avenue detectors would have begun to sort itself in anticipation of the exit ramp at Harlem. Certainly this would seem to be so during the peak periods when lane changing would not be easy.

Presumably, then, most of the exit ramp volume should be in the shoulder lane, less in the middle lane, and a negligible amount in the median lane at East Avenue. This conviction led to the adoption of the following form for the prediction of RV_o:

$$(RV_o)_{t+1} = P_1X_t + P_2Y_t + P_3Z_t + P_4R_t \qquad \text{(B-7)}$$

in which

X_t = volume count on lane 1 at East Avenue at time t;
Y_t = volume count on lane 2 at East Avenue at time t; and
Z_t = volume count on lane 3 at East Avenue at time t.

The last is added to permit some reflection of continuity of flow based on the previous minute's count.

A curve of the form of Eq. B-7 fitted to the four hours of data, using the method of least squares, was evaluated as

$$(RV_o)_t = 0.30425\,X_t + 0.19348\,Y_t + 0.17180\,Z_t - 0.00894\,R_t \qquad \text{(B-8)}$$

For this expression, the sum of the squares divided by the number of data points was computed to be 12.27; for the case where the average value is used for ramp volume, 11.96. Later it was decided not to use this form for predicting RV_o. First, it did not appear that the errors associated with adopting an average value would be any worse than those for the fitted curve. Second, the use of coefficients derived from fitting a curve to one day's data would not be appropriate to a calculation for every day. Third, the "feel" that ramp volumes might be a function of lane distribution at the region just upstream of the ramp could not be extended to lane distributions well upstream of the ramp. Therefore, the prediction of exit ramp volume was included in the calculation of V_o, as discussed in the following.

Having suggested the means for predicting volume in at East Avenue, V_i, and considered one for RV_o, the final piece of data needed, mainline volume out at Harlem, V_o, may be discussed. The logic used to determine predicted values for V_o is based on an assumption for which no data are available to confirm or deny its use. It is hypothesized that traffic performing at a point (maintaining a volume flow) under the volume density curve can continue to maintain this volume count as density increases until such time as the performance point falls on the curve. This is shown in Figure B-2, from which it can be seen that the volume can vary in an infinite number of ways as density

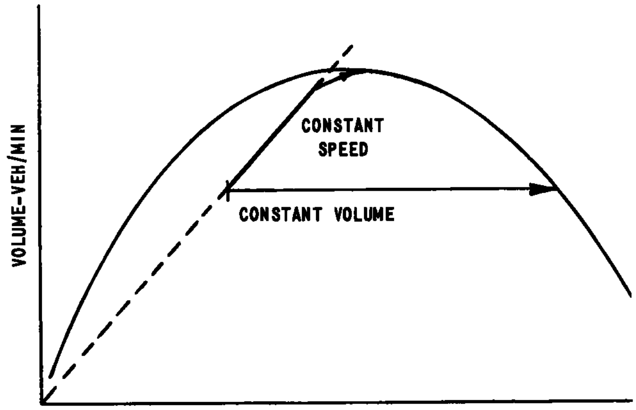


Figure B-2. Volume-density curve, showing constant volume and constant speed relationships.

increases, so long as it stays below or on the curve. The curve itself represents the maximum values that will be developed when the drivers mandate the minimum headways that they will accept as a function of speed. Maintaining constant volume as density increases implies a linear decrease in speed. From the standpoint of driver psychology, it might be more reasonable to expect that the drivers will maintain constant speed until the minimum headway for that speed is approached, after which the speed (and volume) will be reduced. Such a line also is shown in the figure. In the computations performed the constant volume line was used.

Computational Procedure

Four hours of data taken on May 8, 1963, from 3:00 PM until 7:00 PM are used for the computation. The logic used is that given in the earlier report (*1*, pp. 8-9). The procedure is as follows:

1. Compute initial density, ρ_t, at $t = 0$ from Eq. B-1.
2. Compute $(v_i)_{t+1}$ from V_i, the summed lane volumes at East Avenue.
3. Compute $(v_o)_{t+1}$ from

$$(v_o)_{t+1} = 1.1137 \text{ (Channel 27)} + 1.96442 \text{ (Channel 49)} + \text{Channel 29} \quad \text{(B-9)}$$

4. Compute density, ρ_{t+1}, from

$$\rho_{t+1} = \rho_t + \left[\frac{\text{Item 2} - \text{Item 3}}{2.55}\right] \quad \text{(B-10)}$$

5. Test if $\rho_{t+1} > \rho_t$. If no, set the values of ρ_{t+1} and $t + 1$ of Item 4 into Item 1 and continue at Item 2. If yes, set the values of ρ_{t+1} and $t + 1$ of Item 4 into Items 1 and 6 and continue below.
6. Compute predicted densities for 1, 2, 3, 4 and 5 min in the future from t, as follows:

$$\rho_{t+\Delta t} = \rho_t + \Delta\rho_{(t \text{ to } t + \Delta t)} \quad \text{(B-11)}$$

and

$$\Delta\rho = \frac{\Delta t\,(v_i - v_o)_{t + (j+1)\Delta t}}{\tau\, nl} \quad \text{(B-12)}$$

in which $\Delta\rho$ is for the period t to $t + (j + 1)\Delta t$ as j goes from 0 to 19 in each successive calculation; $t = 15$ sec (0.25 min); $\tau = 1$ min; $nl = 2.55$ lane miles; $v_i =$ (Item 2) $+ \frac{dv_i}{dt}\frac{\Delta t}{\tau}$ for $\frac{dv_i}{dt}$ at $t + j\Delta t$; $\frac{dv_i}{dt}$ is from stored data (see Fig. B-3); $v_o =$ (Item 3) until $\rho = \rho_{crit}$, after which v_o is obtained by integration under curve of critical $\frac{dv}{d\rho}$ (see Fig. B-4). Stop computation if $\rho \geqq 100$. Print out ρ for $t + 1$, $t + 2$, $t + 3$, $t + 4$, and $t + 5$. Then return to Item 2.

Results of Computations

The four hours of data that were received were processed on an IBM-704 computer in accordance with the procedure described in the preceding section. Several sample calculations are shown here followed by a general discussion of the results.

The initial density in the test segment is computed as

$$\rho_t = \text{(Average "point" density at DesPlaines)}_{t+1} - \text{(Density due to Harlem on-ramp volume)}_{t+1} + \text{(Density due to Harlem off-ramp volume)}_t$$

$$= \frac{60}{3}\left(\frac{20.7}{52.23} + \frac{17.24}{46.75} + \frac{13.11}{48.05}\right) - \frac{10.62}{3 \times 0.55} + \frac{13.02}{3 \times 0.85}$$

$$= 18.99 \text{ veh/lane/mile.}$$

The fraction terms inside the parenthesis are the individual lane volumes in vehicles per minute divided by the lane speed in miles per hour. In each case the readings are averaged over the previous 45 sec.

The volume of vehicles into the test segment at 3:01 PM was $(v_i)_{t+1} = 30.53 + 32.42 + 20.06 = 83.01$.

The volume of vehicles out of the test segment at 3:01 PM was $(v_o)_{t+1} = 1.1137 \times 27.71 + 1.96442 \times 14.52 + 14.63 = 74.08$.

Because there are more vehicles in the test segment than at the previous minute, the prediction computation is exercised. First, the density at 3:01 PM is $\rho = 18.99 + (83.01 - 74.08)/2.55 = 22.49$ veh/lane/mile. At a density of 22.49 the maximum volume is 82 vpm and demand is limited. Therefore, the value 74.08 may be used for prediction of V_o for the next time interval. The count into the segment for the time interval (1 min) will be 83.01 increased by an increment derived from the historically observed rate of 1,450 veh/hr². Thus, $V_i = 83.01 + 1450/3600 = 83.41$ veh and the density predicted for 3:02 PM is $\rho = 22.49 + (83.41 - 74.08)/2.55 = 26.15$ veh/lane/mile.

Calculations for the succeeding minutes would be done in a similar manner. Computation of initial density is not repeated. As shown in Figure B-4, the value of V_o would be held constant at 74.08 until density reached the value at which the 74.08 ordinate would fall on the curve, after which V_o would be computed as shown.

The results from these four hours of data were set out as follows:

1. Time.
2. Density at the time, in vehicles/lane/mile.
3. Change in density during the previous minute.
4. Predicted density for time plus 1 min.
5. Predicted density for time plus 2 min.
6. Predicted density for time plus 3 min.
7. Predicted density for time plus 4 min.
8. Predicted density for time plus 5 min.

Initial density was computed to be 19 veh/lane/mile at 3:00 PM. Density increased irregularly until at 4:37 PM the jam density of 80 veh/lane/mile was reached, then continued to rise irregularly until the peak value, 125, was computed for 5:22 PM. From that time on, density declined fairly steadily to a value of 79 at 6:00 PM and continued to decrease, but irregularly, to 12 veh/lane/mile at 7:00 PM. Only one unlikely situation was computed to occur, density went to zero at 3:06 PM.

Because the peak computed density of 125 veh/lane/mile seemed to be inordinately high, it was decided to check this point against measured density and occupancy.

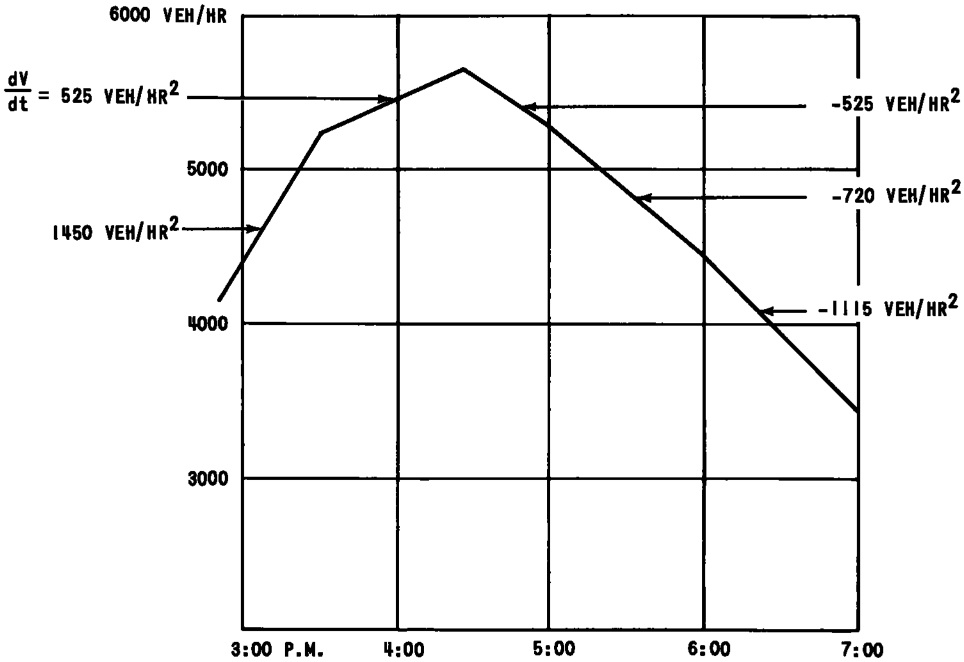
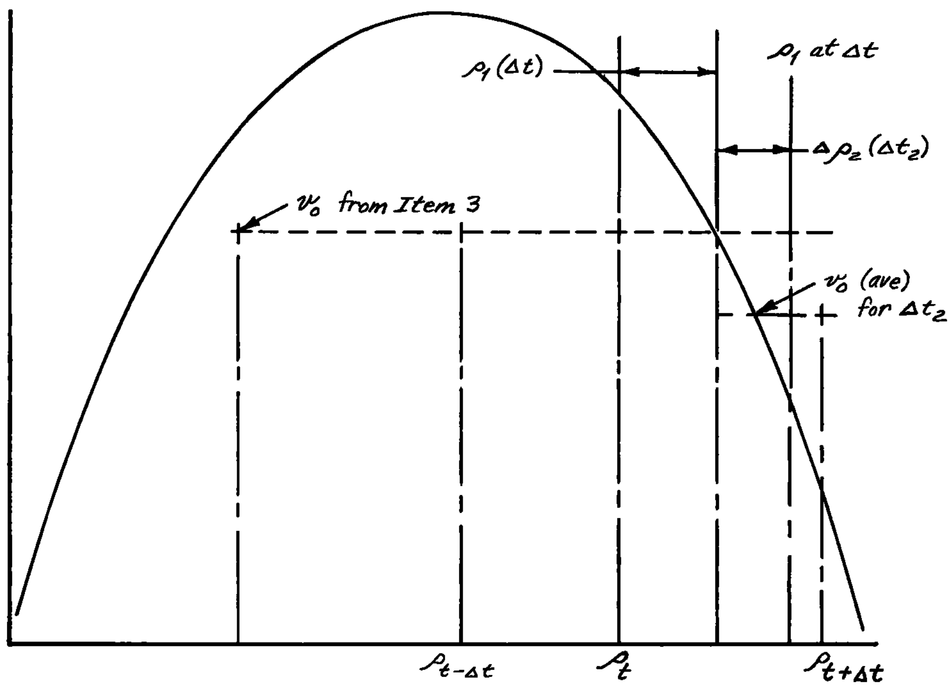


Figure B-3. Hourly volume variations on westbound Congress Street Expressway at East Avenue, October 4-24, 1962.



a. Compute ρ_1 at $t+\Delta t$ as if v_0 persisted at the constant level.

b. Determine average v_0 for the interval Δt_2 .

c. Use average v_0 for the computation of actual $\Delta\rho$ during the period from Δt_1 to Δt .

Figure B-4. Volume-density curve, showing relationships involved.

At Harlem raw data were available from two lanes. Occupancy on the median lane was measured at 23%, the average for the 45 sec ending at 5:22 PM. Assuming an average vehicle length of 16.2 ft, this occupancy translates to a density of 76 veh/lane/mile. Speed and volume in lane 2 were 30.31 mph and 20.69 vpm, respectively. The calculation of density from these values gives 41 veh/lane/mile. Although it is understandable that density on the two different lanes can be quite different, it is equally evident that the average of the two (60) is less than one-half the density computed from the volume counts, the 125 previously noted. For the three lanes at East Avenue, the six occupancy readings for 5:21 PM and 5:22 PM ranged from 14.69% to 18.53%, with the average at about 17.5%, an indication of about 56 veh/lane/mile, which is in reasonable agreement with the data for Harlem. The data were scanned to determine peak occupancy at East Avenue, which was 30.49% at 4:56 PM on the median lane. This was indicative of a density of 99 veh/lane/mile. The nearest speed detector, at Harlem, was showing about 24 mph at this time, which could hardly be called a "stall" speed.

The spot check was not expanded to all four hours for several reasons, the principal one being that it had become evident that the instrumentation collecting and recording the data included systemic inaccuracies that made the data incompatible with the computational logic. The types of error contained in the data were not made apparent until after some processing had been performed. Several of these are discussed in the following.

Current experience with data taken by equipment identical to that installed on the Congress Street Expressway showed that a set of three volume detectors taking counts at one section on a freeway could vary by more than 2% from another set of detectors taking counts at another section. The significance of this error was made evident when, in a running computation of density as was performed with the Congress Street data, it was found that negative densities were indicated in less than 30 min of "bookkeeping" from a known density. This knowledge prompted a calculation to check the consistency of the counts at East Avenue and DesPlaines, the only two sections at which volume was collected on all lanes. In each of two successive hours the counts showed that 126 vehicles more came out of the road segment than went into it. Although the initial density was not known for certain, the length of road in question and the approximate initial density derived from the occupancy reading provided enough of a basis to make the estimate that a negative density would result before the end of a 2-hr period.

Further examination of the data revealed that volume counts taken in lane 3 at DesPlaines by two different detectors showed differences of as much as 12.6%. This was at 4:40 PM. A query to the Expressway Surveillance Project revealed that the data were analog averaged information, not digital counts. Further, the counts were averaged over a 45-sec period just prior to print-out, which constitutes another possible source of error. As a result of the existence of these systemic errors in the data, it was decided to make no further effort to analyze the data for other possible ways of processing to obtain compatible results. It is expected that the computational logic suggested here can be effective only if volume is sensed and recorded almost without error. The admission of an "almost" gives recognition to the fact that a bias error can be acceptable if, on a long-term basis, the amount of the error is the same at all detector stations. Such a situation might be expected to result from the missing of counts by vehicles passing between lanes as they go through the detectors.

A spot check at Harlem at 4:37 PM, the first time at which "jam" density of 80 veh/lane/mile was computed from the volume data, showed that the occupancy was 18.42%, or 60 veh/lane/mile, and the density computed from the speed-volume detector was 45 veh/lane/mile. At 3:00 PM similar calculations show the density at Harlem to be 25 by either method. Occupancies in the three lanes at East Avenue were indicative of densities of 44, 45, and 33. The value used to start the computational process was 19 veh/lane/mile, as shown at the beginning of this section. It may well be that initial density should have been taken as the average of the occupancy indications at East Avenue, because data on all three lanes were available at that section. This point is now considered to be only of academic interest. From the manner in which the coefficients r_1 and r_2 were determined it is probable that the peak density that would have been computed would be higher than the 125 that was found with the initial density of 19.

At this point the discussion is turned to the matter of the prediction computation. During the course of the four hours for which the data were processed there were 117 min of the 240 for which a forecast of increasing density was made. Of this number, 51 were false alarms, for in the succeeding minute the density was computed to have decreased. However, of the 51 false warnings, only 9 indicated that a jam density would be reached within the next 5 min. Thus, an unwarranted action might have been taken 9 times during the course of the four hours. If, as a precautionary measure, the warning would be sounded only when two successive predictions of jam density were made, the number of false warrants would have been reduced to a single case.

Because of the considerable difference between the computed and observed densities, the actual numbers previously given may not be indicative of the period during which a jam density was experienced. Indeed, the raw data indicate that moderate speeds were maintained on the expressway at a time when measured occupancies clearly indicated that stall conditions should prevail. However, the numbers do indicate that an appreciable fluctuation exists, so that predictions based on long-term averaged volume data are not reliable, especially during periods of low flow. This was expected, once the "scintillation" of volumes at Cicero and Austin were noted in the graph, previously mentioned. Thus, the conclusion can be drawn that the use of historical data for prediction purposes should be supplanted by current data taken upstream. It was not done in the present case for the reason previously noted, sampling the traffic flow in one lane only was too unreliable for estimating volume on all lanes. An observation is made concerning the scintillation of the volume flows at East, Austin and Cicero Avenues. The preliminary graph showed 16 peaks in flow at East Avenue during the hour from 3:00 to 4:00 PM. In this same period there were 12 peaks at Austin, just upstream of East, and 13 peaks at Cicero, which is the first instrumented station upstream. This relative similarity would seem to lend weight to an argument in favor of using all real-time data for compression prediction.

How consistent was the occurrence of compression? Jam density (80 veh/lane/mile or more) was computed to exist from 4:37 until 5:59 PM. During this 82-min period 15 false alarms (density did not increase in the following minute) were computed, just a little less than the average of 21.2% for the 4 hr. However, 27 min of valid compression prediction were computed, so that the number of false alarms was $15/(15+27) = 36\%$ of the forecasts for the 82-min period. The false alarm average for the 4 hr was $51/117 = 43.5\%$ of the forecasts, and that for the period when the density was computed to be below jam density was $(51-15)/(117-42) = 48\%$ of the forecasts.

Although this sample of one case, along with all of its shortcomings discussed previously, leaves much to be desired in the way of a level of confidence, it is this indication of the behavior of traffic at high volumes that is being sought. Another indication of the persistence of the flow tendency was noted—that the longest verified prediction during the high-volume period was 10 min of increasing density, whereas during the remainder of the time the tendency persisted for only 4 min.

Concluding Remarks

The limited amount of work that has been done in processing the data taken by the Expressway Surveillance Project shows that no command and control scheme that depends on serial computations with summed data can function with the type of data acquisition and recording system now being used. Systemic errors exist, not only in the acquisition of the information but also in the fact that analog, rather than digital, recording used in the equipment leads to sizeable errors. The equipment is adequate for the purpose of making traffic surveys.

However, the foregoing remark should not be misconstrued to mean that effective controls cannot be devised that will work with the present equipment. Some thought should be given to other logic processes that may be compatible with the existing instrumentation.

The large variations that are possible in the volume-density maximum performance curve lead to the impression that any congestion prediction system might do well to monitor the speeds of the vehicles rather than to depend on the selection of the proper curve for the particular site and operating conditions for a segment of freeway. This is not meant to imply that the curve is not to be used, only that it should not be the sole basis for the effecting of controls.

The data have, however, furnished some slight evidence that the compression process toward congestion, barring an accident, may persist for a long enough time so that it may be sensed and remedial measures put into effect.

SECOND APPLICATION

This section presents the results of an application of data taken on the John Lodge Freeway in Detroit to a computational logic designed to predict the onset of congestion of traffic on a segment of the freeway. The principal reason for using this second set of highway data to test the logic was that the experiment on the John Lodge Freeway originally was designed as a real-time operation and called for data handling different from that for the Congress Street computation. The latter was done on an *a posteriori* basis and there was no visual confirmation of the traffic situation on the expressway. However, the data available for Congress Street included some information on traffic conditions upstream of the test segment and also some historical performance on the expressway at the beginning of the test segment.

Source of the Data

The test site was a 2,217-ft length of the John Lodge Freeway located between the detectors at Hamilton and Calvert Streets in the northwest-bound lanes. This direction of the freeway carries peak traffic during the afternoon hours from 3:00 PM to 6:00 PM. Within this length there is a small amount of curvature and an access ramp joins the three lanes just short of the end of the segment at Calvert Street.

The test site was selected after an aerial observation of traffic flow in it during the peak period. Visual appraisal of the traffic indicated that density was sufficiently uniform at high flow rates so that minor variations could be expected in a half-mile length but not enough to disprove the applicability of the computational method. A shorter length was rejected as being inappropriate for a calculation interval of 1 min, inasmuch as many of the vehicles would pass completely through the test segment and be well downstream during the 1-min interval. A length of freeway of nearly one mile might have served better, but this could not be obtained except with some undesirable changes in geometry.

A sonic presence detector was located over each lane at the two end points. Counts for each detector were totaled for each minute and printed on paper tape. Data collection took place at the Control Center, where were located the printer-counters and all the video monitors and signal control equipment. Individual vehicles picked at random during each minute were timed through an 850-ft length of roadway. This was done manually by observer personnel of the Control Center. In addition, they counted the numbers of vehicles entering the freeway from the access ramp and also were able to determine the exact number of vehicles in the test segment at the beginning of each half hour during the data collection period. Actually, it would have been possible to maintain a precise density check by the method of repeatedly counting the vehicles entering the test segment, starting at the beginning of each minute. However, this arrangement would not lend itself to automatic computation. The manner in which the data were processed is discussed in the following.

Technical Approach

Performance curves have been generated by several investigators for the number of vehicles passing a point as a function of speed. These curves are based on the results of many observations of real traffic systems and show, therefore, envelopes of maximum performance for representative mixes of vehicles and drivers. Although the general shape of the curves is similar for nearly all situations that were sampled, the numbers may vary to reflect characteristics that are unique to each site. A set of curves from data taken on an urban freeway shows that peak volume occurs at about 38 to 40 mph and that volume degenerates significantly with speed variations on either side of the maximum. Therefore, it should be a simple matter to determine lane density at practically stalled flow. This average density is what is used as a criterion for congestion.

The volume-density curve used for these calculations is shown in Figure B-5. When the data from both sources were analyzed it was found that the volumes at high densities were much higher than shown by the curve. The difference may account for the fact that jam conditions were not experienced at some times when the computed densities indicated that traffic should be nearly at a halt.

The scheme for predicting the onset of congestion is

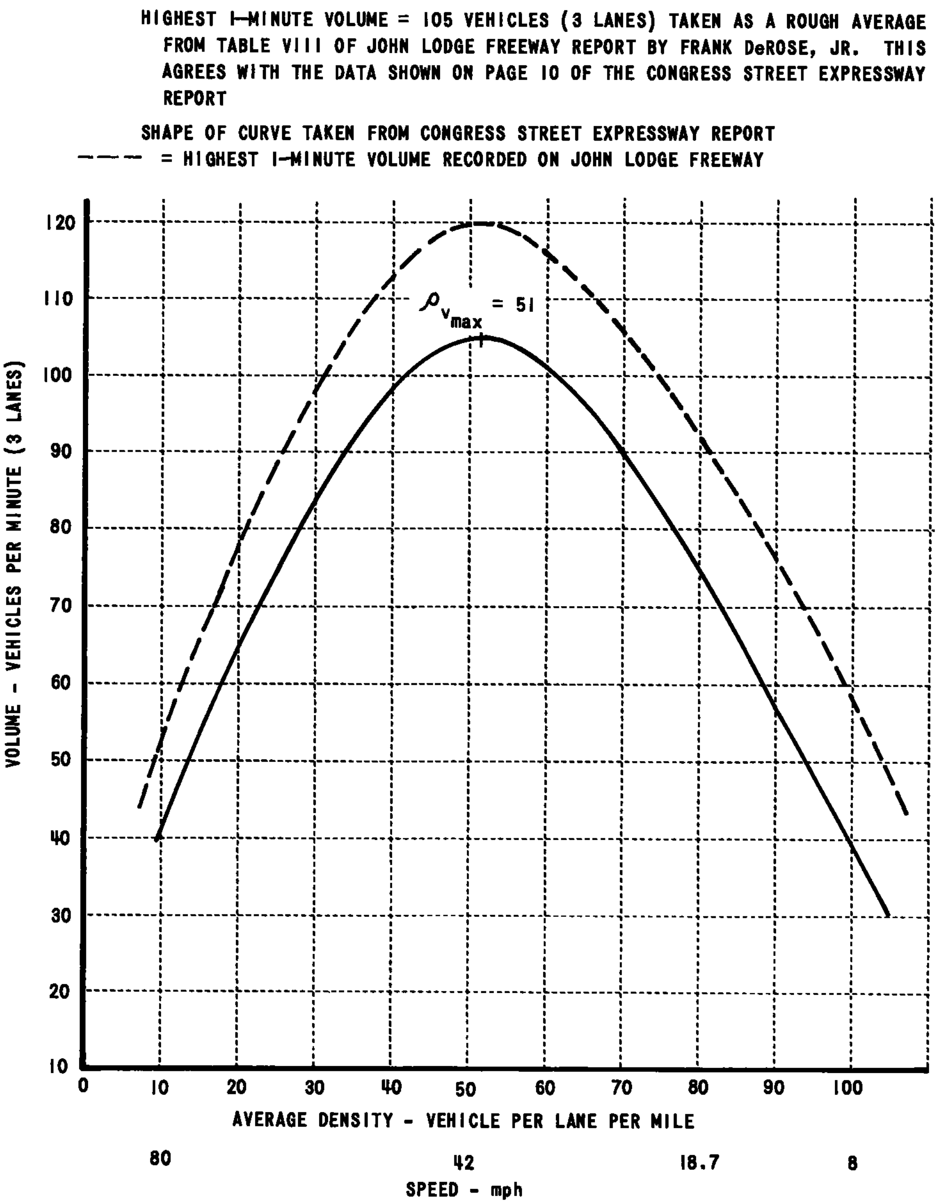


Figure B-5. Volume vs average density.

based on the expectation that traffic will perform somewhat in keeping with the volume-density curve. It is possible for traffic to perform at points below the curve, but this situation would be expected only when the traffic demand is low. The latter might be said also for traffic performing along the curve at densities below that for maximum volume. In the computational procedure a continuous minute-by-minute count of vehicles moving into and out of the test segment is used to update the average density in the segment. Initial value of the density at the beginning of the run is determined by a visual count of the vehicles in the segment. This was made possible through the use of the video presentation of the test site.

In the computational logic it was assumed that whenever an increase in density was computed for a given minute the flow of vehicles into the segment, v_i, would remain constant over the period of prediction. Some assumption had to be made because there was no instrumentation upstream of the test site that would give data on anticipated flow.

Vehicle flow out of the segment, v_o, is based on the assumption that during the period of density prediction v_o will remain constant until the density increases to the point at which v_o falls on the solid curve of Figure B-5. After that v_o will follow along the curve as density rises. Additional discussion of this assumption is given under "First Application."

The implications of the foregoing means for determining v_o for the period of the density prediction computation are that density rise rate will remain constant until v_o falls on the descending portion of the volume-density curve, after which density will rise increasingly until traffic comes to a virtual halt.

Because the congestion prediction experiment originally was planned for computation in real time, the logic used was worked up into overlay charts rather than being programed for machine computation. The logic is that used in the previous section on "First Application." The basic density-time grid is shown in Figure B-6. The heavy solid rays represent the overlay of the linear extrapolation of density until v_o falls on the volume-density curve. The dashed lines represent an overlay for the extrapolation of density over the period when v_o is a function of density. Several of the second type of overlay were needed because of the nonlinear behavior of the volume-density relationship, each overlay being a family of v_i curves for a given v_o.

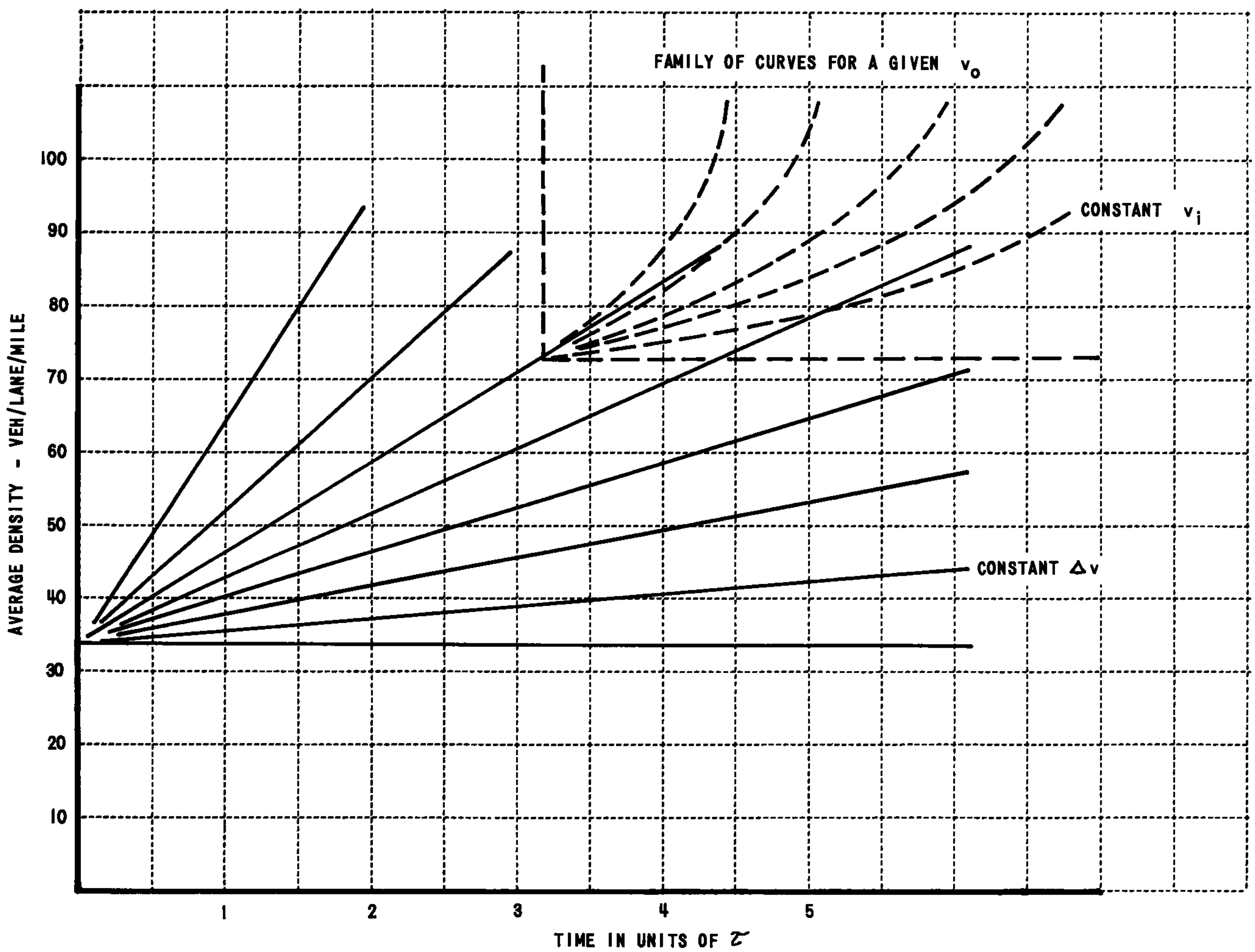


Figure B-6.

OCTOBER 8, 1963

VOLUME DATA TAKEN ON JOHN LODGE FREEWAY

TIME	HAMILTON V_I	CALVERT RAMP V_O	ΔV	COMMENTS
3:0				INITIAL COUNT IS 41 VEHICLES
+1	83	2 81	4	
+2	60	2 74	-12	
+3	65	3 67	1	
+4	94	5 94	5	
+5	76	4 86	-6	
+6	93	2 85	10	
+7	90	3 102	-9	
+8	87	2 99	-10	
+9	89	1 95	-7	
+10	89	6 103	-8	
+11	92	4 87	9	
+12	94	4 98	0	
+13	109	1 86	24	
+14	80	4 95	-11	
+15	92	1 85	8	
+16	95	5 100	0	
+17	78	5 104	-21	
+18	96	5 100	1	
+19	88	1 76	13	
+20	66	3 99	-30	
+21	83	2 99	-14	
+22	91	2 104	-11	
+23	90	3 101	-8	
+24	90	5 88	7	
+25	84	6 107	-17	
+26	93	4 91	6	
+27	84	4 101	-16	
+28	103	4 105	2	
+29	96	3 112	-13	
+30	75	5 100	-20	

SUM OF POSITIVE ΔV = 90
SUM OF NEGATIVE ΔV = 213
DENSITY WOULD GO NEGATIVE

Figure B-7.

OCTOBER 8, 1963

CORRECTED VOLUME DATA - JOHN LODGE FREEWAY

TIME	HAMILTON V_I	CALVERT V_O	ΔV	COMMENTS
3:00				INITIAL COUNT IS 41 VEHICLES
+1	87.5	79	8.5	
+2	63.2	72	-8.8	
+3	68.5	64	4.5	
+4	99.0	89	10.0	
+5	80.1	82	-1.9	
+6	98.0	83	15.0	
+7	94.8	99	-4.2	
+8	91.7	97	-5.3	
+9	93.8	94	-0.2	
+10	93.8	97	-3.2	
+11	97.0	83	14.0	
+12	99.0	94	5.0	
+13	114.9	85	29.9	
+14	84.3	91	-6.7	
+15	97.0	84	13.0	
+16	100.0	95	5.0	
+17	82.2	99	-16.8	
+18	101.0	95	6.0	
+19	92.8	75	17.8	
+20	69.5	96	-26.5	
+21	87.5	97	-9.5	
+22	95.9	102	-6.1	
+23	94.8	98	-3.2	
+24	94.8	83	11.8	
+25	88.5	101	-12.5	
+26	98.0	87	11.0	
+27	85.4	97	-11.6	
+28	108.4	101	7.4	
+29	101.0	109	-8.0	
+30	79.0	95	-16.0	

Figure B-8.

OCTOBER 8, 1963

JOHN LODGE FREEWAY

TIME	P_{t-T}	V_I	V_o	Δ P	P_t	OCC	SPEED	P_{t+1}	P_{t+2}	P_{t+3}	P_I = 95 T TO P_I	COMMENTS
3:0 +0					n/1.26 32.5		579/TIME					
+1	Pt-T	87.5	79	6.8	39.3	stoppage	50 55.1	46	53	59	6.2	ORDINARILY WOULD NOT PREDICT
+2			—	7.0	32.3		42.0					
+3		68.5	64	3.6	35.9		51.2 48.3	39.5	43	46.5		
+4		99	89	7.9	43.8		53.6 51.6	51.5	59.5	67.5	6.6	
+5			—	1.5	42.3		56.2 51.2					
+6		98	83	11.9	54.2		53.6 46	66	78	95	3.0	
+7			—	3.3	50.9		49 39.1					
+8			—	4.2	46.7		46.7					
+9			—	.2	46.5		48.2 43.2					
+10			—	2.5	44.0		52.6 51.6 43.9					
+11		97	83	11.1	55.1		52.6	66	77	97	3.0	
+12		99	94	4.0	59.1		50 52.6	63	67	72	4.5	
+13		114	85	23.7	82.8		51.6 46.7	106	JAM CONDITION SHOULD EXIST			
+14			—	5.3	77.5		41 39.6					
+15		97	84	10.3	87.8	SEVERE SLOW-DOWN	28.6	98	JAM CONDITION SHOULD EXIST			
+16		100	95	4.0	91.8	"	21.9		"	"	"	"
+17			—	13.3	78.5	"	30.3					
+18		101	95	4.8	83.3	"	31.1 27	88	100	JAM CONDITION SHOULD EXIST		
+19		92.8	75	14.1	97.4	STOPPAGE ALL LANES	25	JAM CONDITION EXISTS				
+20			—	21.0	76.4		19.3					
+21			—	7.5	68.9	8	31.8					
+22			—	4.8	64.1		26.8					
+23			—	2.5	61.6		27.8					
+24		94.8	83	9.4	71.0		30.8	80	98	JAM CONDITION ≈ [illegible]		
+25			—	9.9	61.1		34					
+26		98	87	8.7	69.8		32.2 26.8	80	JAM CONDITION ≈ 1.6			
+27			—	9.2	60.0		38.1					
+28		108.4	101	5.9	66.5		34	70	74.5	82	3.9	
+29			—	6.3	60.2		32.9					
+30			—	12.7	47.5		28.6					

Figure B-9.

Application of the Technique to the Data

The data sent by the Control Center consisted of a set of tapes on which were printed the vehicle counts for each of the three lanes at the Hamilton and Calvert Street detector locations. The print-outs were the counts for each minute between 3:00 PM and 6:00 PM on October 8, 1963. In addition, the number of vehicles in the test segment at the beginning of each half hour during the data recording period was counted and recorded on a sheet containing the count of the number of vehicles entering the freeway each minute from the Chicago Street access ramp. A third set of data sheets contained, for each minute of the data recording period, the time required for at least one vehicle to traverse an 850-ft length of roadway located in the test segment. The time was entered in one of three columns, each corresponding to one of the lanes of the freeway. Also noted on those sheets were visually noticeable reductions in the movement of the traffic, from slowdowns to complete stoppages, and the number or numbers of lanes affected. In the data, lane 1 denotes the median lane.

The first operation was to sum the lane counts at Hamilton and Calvert Streets, respectively, for each minute. These were posted on a sheet with the Chicago access ramp counts. Net change in the number of vehicles in the test segment for each minute was computed by subtracting the ramp count from the Calvert Street count and then subtracting the remainder, V_o, from the Hamilton Street count, V_i. Each sheet covered 30 min of data, and the number of vehicles in the test segment was posted for the beginning of the half hour. It was in the posting of values of $V_i - V_o$ ($= \Delta V$) that the errors in the data were detected. It was noticed that negative values of ΔV were predominant and a check soon showed that within each half hour of the 3 hr for which data were recorded the density within the segment would go negative, which is an impossible situation.

Inasmuch as an actual count of vehicles in the segment was available every half hour, except at the end of the data logging period, it was possible to apply a correction factor to the total half-hour volume for one of the stations in order "to make ends meet." Re-examination of the individual lane counts led to the conclusion that the trouble probably was in the volume count for lane 1 at Hamilton. A correction factor to be applied to the minute counts at Hamilton for each half hour, except for the last, was computed and found to be 1.052, 1.043, 1.047, and 1.059, with an average value of 1.050.

The correction factor for each half hour was applied to the V_i counts within that half hour, except for the last, for which the average value of the five correction factors was used. Values of ΔV then were computed, converted to changes in density by

$$\rho = \frac{\Delta V}{nL} \tag{B-12}$$

in which n is the number of lanes, and L is the length of the test segment, and the changes in density were used to update the initial value of density determined from the vehicle count within the segment.

Actually, for the specific situation shown here, all computations could have been made in terms of volumes and numbers of vehicles. This would have required a change in the abscissa scale of the volume-density curve on which the expected behavior of the traffic was based. However, even though a greater number of computations was entailed, the procedure was retained in its more general form.

For each entry in which an increase in density was computed, there was also computed an extrapolation into the future, for the 3 min following, of the density that might be expected, or it was noted that "jam" density had been reached and when. In view of the experience with the data used in the Congress Street investigation, the density at jam was estimated at 95 veh/lane/mile instead of 80, which had been estimated from the lower curve of Figure B-5. Sample pages of the calculation sheets are shown in Figures B-7, B-8, and B-9.

Results of Computations

Over the 3-hr period during which the data were taken, 94 minute counts indicated an increase in density in the test segment. Of the 94, prediction computations were made for 50. In the other 44 cases either the computed density already was in excess of the jam value or it was so low that, in conjunction with a very low density rise rate, no prediction computation was warranted.

Slowdowns and stoppages were observed by Control Center personnel monitoring the video displays. They recorded occasions during 33 minutes in which traffic flow deterioration was visibly noticeable. To match this, of the 50 prediction computations made, 32 included predictions of reaching jam density within 3 min. However, this excellent matching is not cited as an indication of the accuracy of prediction.

During the first half hour there were 4 false predictions of jam out of 9, although one of these was in error by only 1 min. In the next half hour 6 predictions out of 13 were in error, but 5 of the 6 were made at a time of high density when any instability could cause a tie-up to develop quite quickly. The density during these 5 minutes was 80 veh/lane/mile or more. It is of interest to note that stoppages occurred on single lanes when the average densities were as low as 67 and 75 (Fig. B-10 at 3:58 PM and 5:02 PM, but note that both were stops in one lane only).

In the period from 4:00 to 4:30 PM there were 3 predictions through jam density. In two of these the jam density was not computed to occur although a stoppage was reported in the time bracketed by one of the predictions. The other was in error by 1 min. During the following half hour there were 6 jam predictions, with but one in error.

In the half hour beginning at 5:00 PM none of the 4 anticipated tie-ups was in error. In fact, it was near the beginning of this time period that a stoppage was missed (the one that occurred at an average lane density of 67).

It is obvious that use of the average of the correction factors for application to the final half-hour's data led to considerable error in density computation toward the end of that period. Four predictions of congestion were

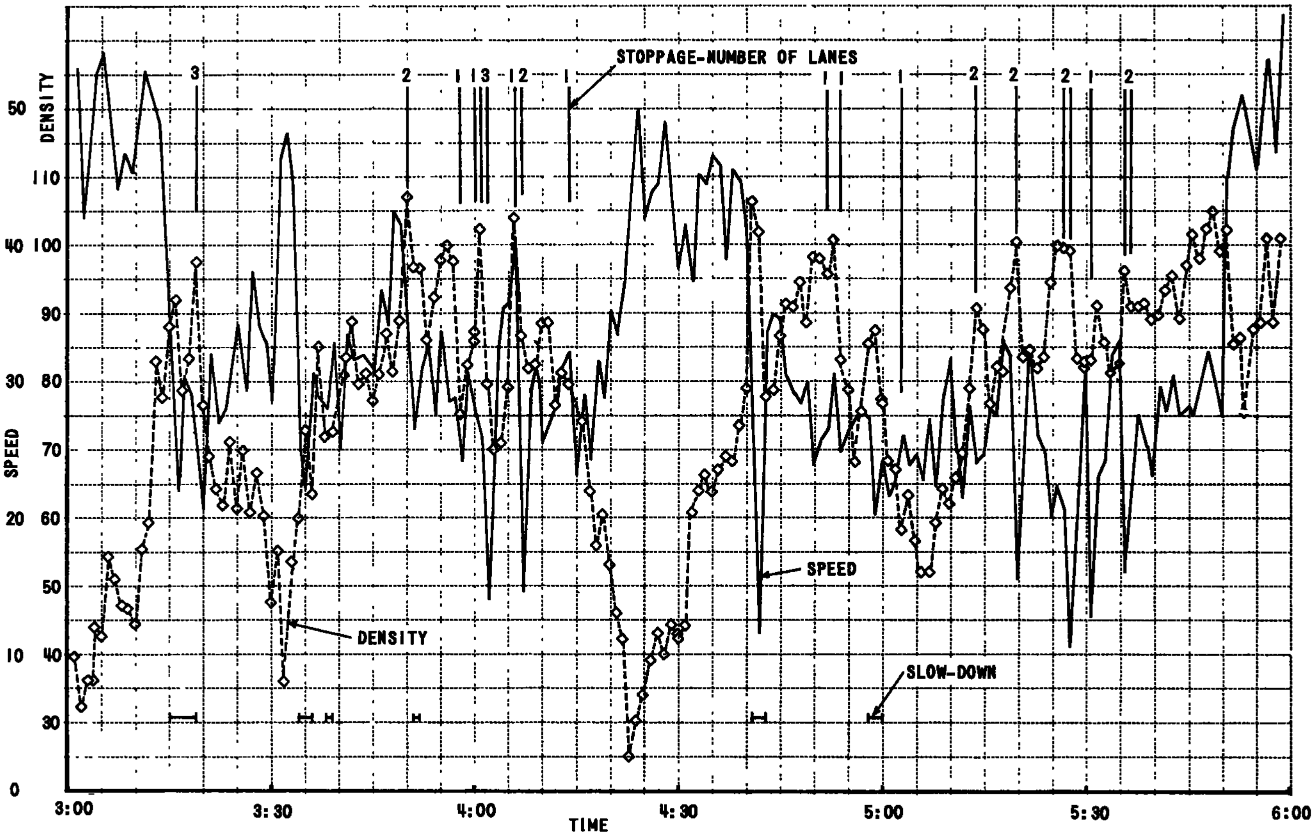


Figure B-10. Speed and density variation, John Lodge Freeway.

made, of which two were false. For the last 15 min of the period the computed densities ran much higher than the observed speeds would indicate were possible. One stoppage was missed altogether; it occurred at a computed density of 83 in the first minute of this period, so the density should be accurate.

Thus, of the 32 predictions made 25 were valid, giving a false alarm rate of about 25%. Two stoppages occurred without warning from the computer. Of the 25 warnings given, only 5 anticipated the congestion by 3 min or more and these were not all continuously confirmed after the initial warning.

The general relationship between speed and density for the 3-hr period is shown in Figure B-10. The expected tendencies are exhibited; low density allows high speed and high density produces slowdowns and stoppages. The general exception is seen in the last 10 min, where, as previously pointed out, it is thought that the correction factor used was too high.

The 5 predictions of jam density that were in error in the half hour from 3:30 to 4:00 PM occurred between 3:40 and 3:46 PM. Examination of Figure B-10 shows that this comparatively high level of both speed and density occurs only at this point during this period. At any other period where densities rise to or above this level, stoppages and/or slowdowns are noted to occur.

The occurrence of these inconsistencies is to be expected. Variations in segments of the driver population are bound to occur, as well as the effect of perturbations in the traffic stream. Traffic characteristics derived from the statistics from macroscopic observations may be used for generalized studies, but the kinds of departures that happened from 3:40 to 3:46 PM must be considered as part of the distribution. The "spatter" of the speed-density behavior of the traffic for this data period also can be seen from the ill-defined tendency in Figure B-11.

Conclusions and Recommendations

The conclusions that may be drawn from the foregoing treatment of the data taken on the John C. Lodge Freeway are much the same as those concerning the processing of data from the Congress Street Expressway with a somewhat similar logic. These are that the congestion prediction scheme should not use volume counts for a serial bookkeeping of the density on the roadway, instrumentation should be installed over all lanes at about ½-mile intervals in order to sense volumes and speeds, and analog recording of the incoming data should be avoided if at all feasible to do so.

Another approach to this problem probably should make concessions to the order of accuracy that may be expected from current instrumentation or reasonably priced future instrumentation. Such an approach may be dependent solely on current information being gathered at many points along the roadway during the computation periods. Another facet of the problem will be to try to analyze currently gathered data for the purpose of eliciting criteria for the performance that may be expected from the roadway, taking into account its geometry and the environmental conditions, rather than to use averaged historical data that may not apply to the situation at a particular time.

Additional information that might be useful in effecting control on an urban freeway would be the flow conditions on surface arterials in the immediate vicinity of exit ramps, because it is likely that during periods of high demand stoppages occurring on the streets can be transmitted quickly upstream.

One thing was surely apparent from the processing of the data. Stoppages can occur at relatively moderate densities if an accident or other unforeseen incident occurs. No warning system is envisioned that can predict such a situation, but this is no reason to hold it in low regard if that is its only failing.

In any future development of a data processing scheme it may be desirable to examine more than the first derivative of the varying traffic behavior. This sort of computation may be necessary if warning time is not consistent or is too short. The availability of additional data from detectors upstream may alleviate the latter difficulty.

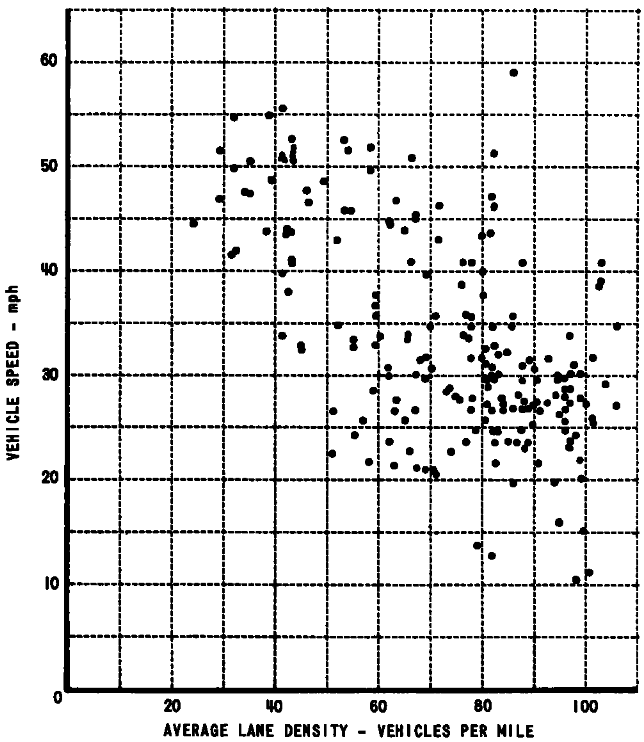


Figure B-11. Speed vs density, at 1-min intervals, John Lodge Freeway, October 8, 1963, 3:00 to 6:00 p.m.

2. AUTOMATIC DETECTION OF STOPPED VEHICLES BY ELECTRONIC MEANS

A brief investigation has been conducted of electronic techniques that could be used to detect stopped vehicles on a freeway. The following assumptions were made:

1. The section of freeway to be considered is essentially straight and free of hills that might obscure a car.
2. The sensor can be located very near to or over the highway.
3. The sensing and reporting system must be entirely automatic.
4. The total width of the freeway, including the center mall, is 150 ft.

A number of sensors (radar, optical, infrared (IR), television (TV), etc.) might be considered for this application. The optical sensors would have little value during periods of fog or heavy smog; the radar would be usable (depending on the operating frequency) except under violent rainfall conditions.

Detecting the presence of a stationary vehicle is much more difficult than detecting a moving vehicle. A number of techniques based on the Doppler effect can be used to detect moving targets, but it is difficult to determine when one additional stationary target has been added to the normal complex of fixed targets such as posts, poles, signs, and terrain discontinuities. This is further complicated by the fact that the freeway subtends a smaller angle with respect to the sensor as the distance increases. For example, at a range of 1,000 ft the freeway subtends an angle of about 9°, whereas at a range of 10,000 ft (about 2 miles) the angle is less than 1°. To obtain proper coverage at short ranges, it will be necessary to utilize a sensor with a wide field of view or scan a narrow-beam sensor. In any case, the necessity for wide-angle coverage at short ranges implies that at maximum range the sensor will be viewing primarily the terrain around the freeway. Thus, in a typical installation there will be a large number of extraneous objects that must be processed by any automatic system.

With the foregoing restrictions in mind, several possible techniques to detect a stationary vehicle are considered. The first utilizes a radar with a moving target indication (MTI) capability. This radar would have an antenna with a narrow beamwidth (1° or so), and the antenna would be scanned through a sufficient angle to view the freeway near the radar. If all cars were restricted from the freeway, the radar presentation would indicate the fixed targets in the field of view. This presentation would be photographed to serve as a reference. In addition, both the photograph and the radar indicator would be masked to restrict the field of view to the freeway itself. During actual operation the radar would observe these fixed targets plus the moving automobiles and any stationary vehicles. By comparing the radar presentation with the previously photographed presentation, all fixed targets would be eliminated and the only remaining targets would be moving and stalled automobiles. The MTI feature of the radar would identify the moving vehicles and the remaining targets would be stalled vehicles.

A radar system to achieve this function would be expensive. The cost of a single radar unit for air traffic control use is in the vicinity of $100,000. A smaller, lower-powered radar would suffice for freeway use; however, the requirement for the data processor described in the foregoing would more than compensate for any cost reduction achieved by the reduction in size.

Another scheme for detecting stalled vehicles would make use of either a radar or a TV system. First, it would be necessary to obtain a view of the freeway without traffic. In operation, the radar or TV system would scan the desired area and the resultant information would be stored on a storage tube. Periodically, perhaps once every second, the scene on the storage tube would be compared with the previously obtained nontraffic view of the freeway. By comparing these two presentations, the fixed targets can be cancelled out. Because the moving targets are at various positions during each comparison period, their signals will assume an average intensity. The stopped vehicles will remain at the same position during every sampling period and the return from these vehicles will integrate to provide a recognized signal.

This would also be an expensive scheme because of the data-processing circuitry required. The TV system probably would be less costly than the radar, but would have the disadvantages of less resolution, possible degradation due to weather, etc.

The schemes described in the foregoing are based on radar and data processing equipment which has been built and tested. It might be possible to consider other detection means if more basic data were available. For example, what is the temperature, as measured by an infrared detector, of a stalled vehicle with respect to its surroundings or to moving vehicles? There may be a signature here which may simplify the detection problem.

In conclusion, it can be stated that well-known electronic techniques are available to detect the presence of a stalled vehicle on a freeway. It would be very expensive, however, to develop and build equipment incorporating these techniques. Emphasis should be placed on investigation of techniques that will permit use of simpler and less expensive equipment to accomplish this task.

3. ACOUSTIC AND RADIO-FREQUENCY PRESENCE DETECTOR LIMITATIONS

This section discusses some basic physical properties that govern the operating characteristics of acoustic and radio-frequency traffic detectors, which depend on active radiation of a signal and reception and processing of the echoed return.

PRESENCE DETECTORS

The presence detectors discussed here are the over-the-roadway type. These instruments radiate a pulse vertically downward to the pavement and receive the reflected signal. The determination of "vehicle" or "no vehicle" is based on the time elapsed for return of the pulse. Theoretically, the signal process is timed as shown in Figure B-12. The detector is installed at a nominal height, h, over the roadway.

Assume that at some time, $t = 0$, the antenna or transducer emits the leading edge of a pulse of width, w, time. Then, as shown, the time required for the pulse to travel to the roadbed and for the reflected signal to be completely returned into the receiver is

$$t = \frac{2h}{V_s} + w \qquad \text{(B-13)}$$

in which V_s is the propagation velocity of the signal. In theory, the maximum rate at which such pulses could be transmitted, or pulse repetition frequency (prf), would be the reciprocal of t. However, practical considerations require that some delay time, d, be included to allow for switching from the receiver to the transmitter circuit. Thus, the real prf would be

$$\text{prf} = \frac{1}{\tau} = \frac{1}{\frac{2h}{V_s} + w + d} \qquad \text{(B-14)}$$

The implications of prf, in a presence detector, are the number of pulses that will "hit" on a vehicle moving under the detector, the higher the prf the greater the number (the closer the spacing) for a given vehicle speed. The greater the number of hits on the vehicle the finer will be the resolution of the time the vehicle spends under the detector.

Theoretically, no matter how high the prf, there may be an error in time measurement that can be a maximum of τ too long or 2τ too short, depending on the method of signal processing. The source of error is shown in Figure B-13 from the following line of reasoning.

Suppose a vehicle is moving under the detector at speed, s. Then, the spacing of signal "hits" on the vehicle will be s/prf. Let it be assumed that in one data processing scheme the time the vehicle spends under the detector is computed by the number of hits, n, divided by the prf. If such a scheme was applied to the case in Figure B-13, where the first and last hits barely generated a return signal, shown by the broken lines, the processor would indicate that the vehicle remained under the detector for a time τ more than it did in fact.

If, on the other hand, the signal is processed by computing the time under the detector to be that period only for which "presence" is sensed, the indication would be that the vehicle was under the detector for a period nearly 2τ less than it really was. The locations of these signal contact (echo) points on the vehicle are shown by the solid vertical lines.

Pulse Repetition Frequency

The preceding discussion shows how the time between signal pulses, τ, affects the accuracy of measurement of time under a detector. Now a comparison is made between these times for acoustic and radio-frequency (r-f) detectors.

Velocity of propagation of sound waves in air is given by

$$a = \sqrt{\gamma R T} \qquad \text{(B-15)}$$

in which T is the absolute air temperature, R is a gas constant, and γ is the specific heat ratio. Assuming that the detectors must operate at temperatures as low as -20 F (440 R), the speed of sound will be 1,029 ft/sec. (This is the lowest velocity anticipated for most U. S. applications and is therefore the most critical condition for accuracy.) If, for instance, the height of the detector over the road is 16 ft, the time required for the signal to traverse the 32-ft distance is $2h/a = 0.0311$ sec. A currently used acoustic detector of this type uses a pulse width of 1 millisecond and if another millisecond is added for switching delay, the time between pulses will be 0.331 sec, or a prf of 30.2 cps.

Electromagnetic radiation speed, c, is 300×10^6 meters/sec. Time for a signal to traverse 32 ft is about 0.032 microsec. Disregarding pulse width and switching delay time, this traverse time could permit a prf of about 32 megacycles. But these other design considerations of the circuitry are critical and require that the prf be limited to about 200 kc. This would give a resolution of time about 6,000 times as fine as that for the acoustic detector. It would appear that in practice a much lower prf than 200 kc would suffice for the needs of the system.

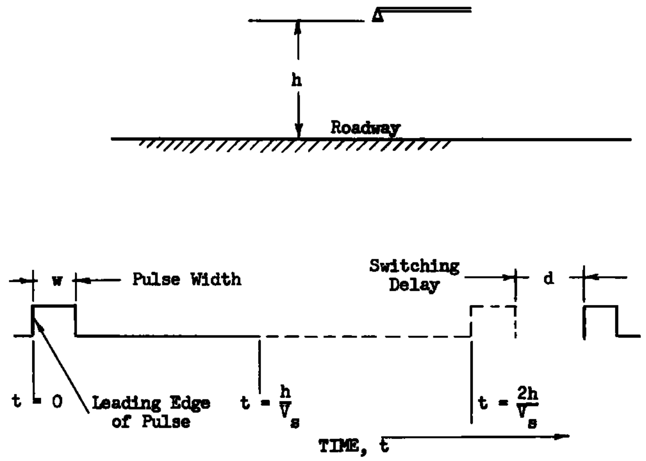


Figure B-12. Signal process timing for over-the-roadway presence detector.

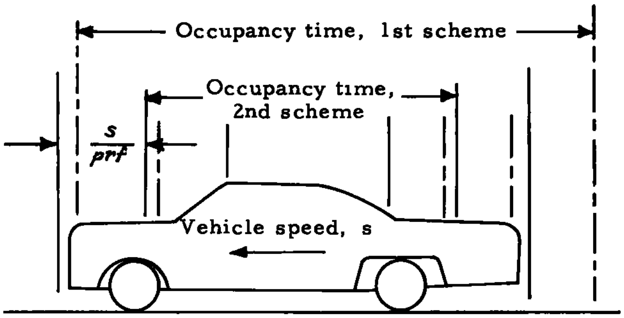


Figure B-13. Occupancy relationship under two detection pulse schemes.

Occupancy Error

The theoretically attainable values of pulse repetition frequency are of academic interest because they point up absolute performance limits. Of more immediate concern are the operating characteristics of detectors in actual use.

One ultra-audio acoustic presence detector uses a prf of 25 cps (or less). A 16-ft long vehicle moving under such a detector at 60 ft/sec (40 mph) would be "hit" every 2.4 ft., giving 6 or 7 echoes, depending on the arrival phase. Depending on the manner of data processing, as shown in Figure B-13, the time under the detector could be computed as $7 \times 0.04 = 0.28$ sec by scheme 1, maximum, or $5 \times 0.04 = 0.20$ sec by scheme 2, minimum. Actual time under the detector is $16/60 = 0.267$ sec. Thus, the error for scheme 1 is 4.9% high and the error for scheme 2 is 25% low. As noted previously, under scheme 1 the length of the vehicle could have varied from just over 14.4 ft to just under 16.8 ft (from 6τ to 7τ in time) for a detector indicated period of 0.28 sec (7τ). Similarly, under scheme 2 the length of the vehicle could vary from just over 12 ft to just under 16.8 ft (5τ to 7τ) for a detector indicated period of 0.20 sec (5τ).

One r-f detector in current use has been limited to a prf of 65 cps for reasons other than the one restricting to a practical upper limit of 200 kc. One of the more pertinent reasons is that it is used in an area having a 30-mph speed limit. In any event, it is seen that this relatively low (for an r-f system) prf has slightly more than 2½ times the resolution of the acoustic detector. For the situation previously described (a 16-ft vehicle traveling at 60 ft/sec), the spacing of the echoes will be 0.924 ft., giving 17 or 18 returns. Time under the detector could be computed as $18/65 = 0.277$ sec by scheme 1, or $16/65 = 0.247$ sec by scheme 2. Thus, the error for scheme 1 is 3.7% high and the error for scheme 2 is 7.5% low. Again, the vehicle could range from just over 15.7 ft to just under 16.6 ft for a detector indicated period of 0.277 sec by scheme 1. It would range from 14.8 ft to 16.6 ft for a detector indicated period of 0.247 sec by scheme 2.

At any given pulse repetition frequency the system errors will increase as vehicle velocity increases or vehicle length decreases.

Concluding Comments

The effect of pulse repetition frequency on accuracy of the measurement of time occupancy has been shown, as well as the derivation of prf from signal propagation speed. The errors that have been computed are systemic to the design philosophy; the assumption has been made that the operation of the equipment is perfect.

The use of time occupancy to convert to speed, based on a statistically determined average length of vehicle, merely compounds the error. It is suggested that these errors may make the use of lane occupancy to compute speed unsuitable for short-term information for a traffic control system.

It should be noted that the introduction of practical considerations of detector design into the discussion has had to do only with the establishment of upper limits on pulse repetition frequency. There are other factors that may be the source of additional errors in either acoustic or r-f systems. One is the finite width of the beam that impinges on either the roadway or the vehicle and another is the strength of the return signal required to enter the receiver in order to register presence of a vehicle. In the latter case it is possible for the shape of the vehicle to attenuate a return signal to the point where it may not emerge from the receiver background noise, thus losing the time for that pulse from the occupancy period.

SENSITIVITY OF SPEED DETERMINATION

Expressions are developed here to give the theoretical speed range over which a vehicle length, L, will be "seen" for a constant time interval, O, by a pulse-type presence detector operating at a given pulse repetition frequency, prf. The limiting conditions are shown in Figure B-14,

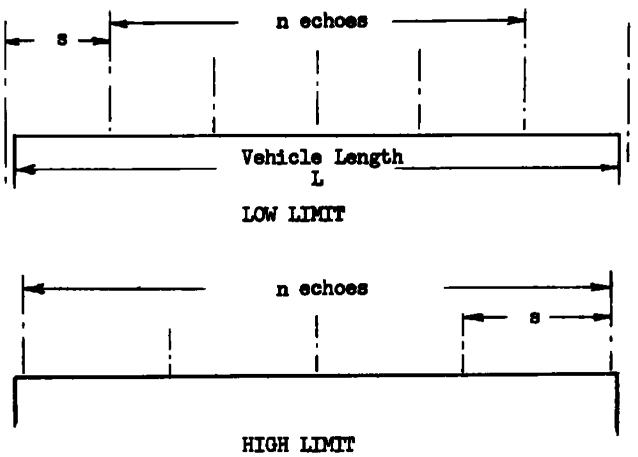


Figure B-14. Maximum and minimum limits for pulse echoes.

which shows that for any vehicle velocity, v, the spacing of the echoes that will be returned from the vehicle will be

$$s = v/\text{prf} \qquad \text{(B-16)}$$

The maximum number of echoes that can be returned from a vehicle of length L moving at velocity v is

$$n = \left(\frac{L}{s} + 1\right) \qquad \text{(B-17)}$$

to the next lower integer. Nominal time occupancy under the detector will be:

$$O = n/\text{prf} \qquad \text{(B-18)}$$

Then, the highest speed, v_{max}, that the vehicle can travel under the detector, for a measured time occupancy of O, is when the number of echoes, n, is barely accommodated on the length of the vehicle. This case is shown as the high limit in Figure B-14. Thus, when

$$s = L/(n\text{-}1) \qquad \text{(B-19)}$$

$$v_{max} = L\ \text{prf}/(n\text{-}1) \qquad \text{(B-20)}$$

Similarly, the lowest speed for the same time occupancy is

$$v_{min} = L\ \text{prf}/(n+1) \qquad \text{(B-21)}$$

To illustrate this variation, a 16-ft long vehicle moving at 60 ft/sec under a detector with a prf of 25 cps would be "hit" every 2.4 ft, giving a maximum of 7 echoes. Thus, $v_{max} = (16 \times 25)/6 = 66.7$ ft/sec, and $v_{min} = (16 \times 25)/8 = 50$ ft/sec.

And what of a distribution of vehicle speeds? If a population of 16-ft long vehicles moves under the detector at speeds distributed symmetrically over the foregoing speed range, the mean speed will be 58.3 ft/sec. The computed speed would be 57.2 ft/sec (assumes $L = 16$ ft).

Minimum and maximum vs computed speeds, over the speed range of interest, are shown in Figure B-15.

Effect of Beam Width on Occupancy Measurement

The foregoing discussion is based on a pulse-type detector having a zero beam angle. Obviously, no such instrument will be found in practice. Therefore, the effect of a finite beam width is investigated.

Suppose the detector emits a signal of half-angle, α, along the direction of travel. It is evident that the entire distance intercepted by this angle cannot be used as a bias to change the value of L in Eq. B-20 and B-21. This point is illustrated in Figure B-16. The solid outline of the vehicle, just tangent to the limit of the beam angle, will not "echo" a return. It is necessary that the vehicle penetrate to some point in the beam, shown by the angle β, so that sufficient echo is received for detection of presence to occur. This is shown by the dotted outline. Then, if the distance from the centerline of the beam to the edge of the vehicle is written as βh, Eq. B-20 and B-21 become

$$v_{max} = \frac{(L + 2\beta h)\ \text{prf}}{n - 1} \qquad \text{(B-22)}$$

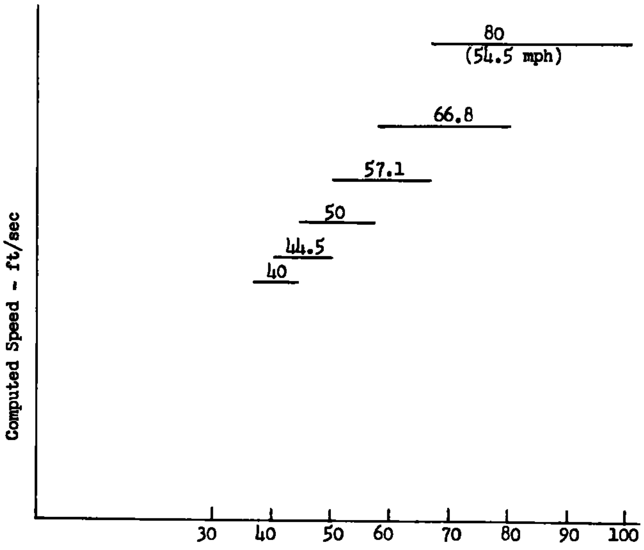


Figure B-15. Computed vs actual speeds.

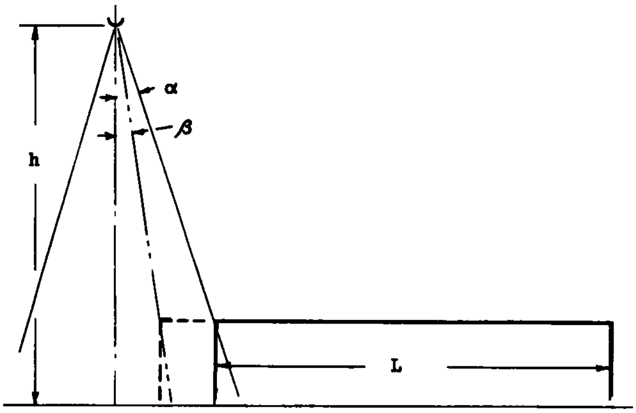


Figure B-16. Effect of beam width on occupancy measurement.

and

$$v_{\min} = \frac{(L + 2\beta h)\ \text{prf}}{n + 1} \qquad \text{(B-23)}$$

This is shown in Figure B-17.

Inspection of Eqs. B-22 and B-23 shows that the change introduced by the quantity βh has a greater effect in the high-speed calculation than in the low.

Concluding Remarks

In this analysis it has been shown that large errors may be expected in the measurement of time occupancy of individual vehicles by low-frequency (20-25 cps) pulse-type presence detectors. Theoretically, a 16-ft long vehicle can pass under a 25-cps detector at any speed from 50 to 66.7 ft/sec and receive and echo 7 pulses from the detector. The nominal time under the detector would be 0.28 sec. Speed, computed from time occupancy and the 16-ft vehicle length, would be 57.2 ft/sec. Thus, the actual speed can be as much as 16.8 percent higher or 12.4 percent lower than the computed speed. The errors are speed related, both absolute and relative, the higher the speed the greater the error. Also, if the actual vehicle length is much different from the assumed value of L, individual computed speeds can have even higher errors than those shown.

The point has been made that a symmetrical distribution of speeds over any of the "steps" shown in Figure B-15 would have an average value different from the speed computed for the nominal time under the detector. The error, for the case cited, would be about 2%, the nominal speed being lower than the average. Of course, in actual traffic there is no reason to suppose that the vehicle speeds will be symmetrically distributed over one of the steps. It would be quite reasonable to expect that a distribution could be biased toward one end or the other. Such a situation could cause significant errors in measuring time occupancies and computing speeds for groups of vehicles, except for one thing.

The preceding commentary presumes that the "hits" are symmetrically distributed over the length of each vehicle. Such a presumption is unwarranted. It is more likely that there will be a considerable randomness in the arrival of successive vehicles, so far as position of the pulses is concerned. Thus, a distribution of speeds at one end of a "step," shown in Figure B-15, would not be expected to be measured in error by the difference between the computed value and the actual mean value. The fact that the randomness of arrivals will cause a change in the number of "hits" will reduce the error.

For instance, referring to Figure B-15, suppose a population of vehicles passes under the detector at speeds over

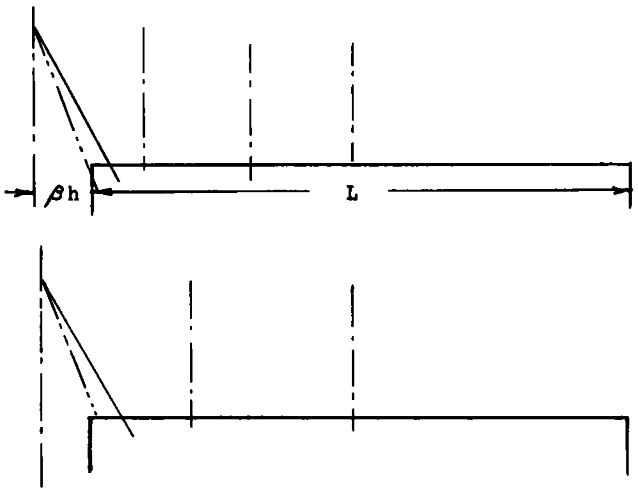


Figure B-17. How beam width affects echo of detector pulse.

the range from 70 to 80 ft/sec. If the arrivals are such that each vehicle produces 6 echoes, the computed speed for the population will be 66.8 ft/sec. If speeds are symmetrically distributed, the actual mean value will be 75 ft/sec, and the error is appreciable. But, in the speed range from 70 to 80 ft/sec the number of echoes can be either 5 or 6, depending on the phase of arrival. If the arrivals are uniformly distributed in speed and phase, the expected number of echoes will be 5.32. Thus, the computed mean speed would be 75 ft/sec, which agrees with the actual mean speed. The errors are introduced when vehicle lengths are assumed (not known) and the speed and phase distributions are not uniform.

Whether or not arrivals will be random in real traffic is a moot point. One *could* imagine some ordered spacing deriving from the mean speed and the speed-headway relationship. Even on a statistical basis, these expected values could produce some bias in the readings for platoons of vehicles. The point is interesting, but academic at this time.

Thus, it is concluded that any presence detector equipment that is limited to pulse or interrogation rates of about 30 cps or less may give satisfactory performance for some traffic control and data acquisition where large sample sizes are used. The equipment is not sensitive or accurate enough for some of the more advanced control techniques that have been suggested for urban freeways and networks. Emphatically, this type of detector is not suitable for measurement of time occupancy (and speed computed therefrom) for purposes of research into traffic behavior.

Published reports of the

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

are available from:

Highway Research Board
National Academy of Sciences
2101 Constitution Avenue
Washington, D.C. 20418

NCHRP Report No.	*Title*
—*	A Critical Review of Literature Treating Methods of Identifying Aggregates Subject to Destructive Volume Change When Frozen in Concrete and a Proposed Program of Research—Intermediate Report (Project 4-3(2)) 81 pp. $1.80
1	Evaluation of Methods of Replacement of Deteriorated Concrete in Structures (Project 6-8) 56 pp. $2.80
2	An Introduction to Guidelines for Satellite Studies of Pavement Performance (Project 1-1) 19 pp. $1.80
2A	Guidelines for Satellite Studies of Pavement Performance 85 pp.+ 9 figs., 26 tables, 4 app. $3.00
3	Improved Criteria for Traffic Signals at Individual Intersections—Interim Report (Project 3-5) 36 pp. $1.60
4	Non-Chemical Methods of Snow and Ice Control on Highway Structures (Project 6-2) 74 pp. $3.20
5	Effects of Different Methods of Stockpiling Aggregates—Interim Report (Project 10-3) 48 pp. $2.00
6	Means of Locating and Communicating with Disabled Vehicles—Interim Report (Project 3-4) 56 pp. $3.20
7	Comparison of Different Methods of Measuring Pavement Condition—Interim Report (Project 1-2) 29 pp. $1.80
8	Synthetic Aggregates for Highway Construction (Project 4-4) 13 pp. $1.00
9	Traffic Surveillance and Means of Communicating with Drivers—Interim Report (Project 3-2) 28 pp. $1.60
10	Theoretical Analysis of Structural Behavior of Road Test Flexible Pavements (Project 1-4) 31 pp. $2.80
11	Effect of Control Devices on Traffic Operations—Interim Report (Project 3-6) 107 pp. $5.80
12	Identification of Aggregates Causing Poor Concrete Performance When Frozen—Interim Report (Project 4-3(1)) 47 pp. $3.00
13	Running Cost of Motor Vehicles as Affected by Highway Design—Interim Report (Project 2-5) 43 pp. $2.80
14	Density and Moisture Content Measurements by Nuclear Methods—Interim Report (Project 10-5) 32 pp. $3.00
15	Identification of Concrete Aggregates Exhibiting Frost Susceptibility—Interim Report (Project 4-3(2)) 66 pp. $4.00
16	Protective Coatings to Prevent Deterioration of Concrete by Deicing Chemicals (Project 6-3) 21 pp. $1.60
17	Development of Guidelines for Practical and Realistic Construction Specifications (Project 10-1) 109 pp. $6.00
18	Community Consequences of Highway Improvement (Project 2-2) 37 pp. $2.80
19	Economical and Effective Deicing Agents for Use on Highway Structures (Project 6-1) 19 pp. $1.20
20	Economic Study of Roadway Lighting (Project 5-4) 77 pp. $3.20
21	Detecting Variations in Load-Carrying Capacity of Flexible Pavements (Project 1-5) 30 pp. $1.40
22	Factors Influencing Flexible Pavement Performance (Project 1-3(2)) 69 pp. $2.60
23	Methods for Reducing Corrosion of Reinforcing Steel (Project 6-4) 22 pp. $1.40
24	Urban Travel Patterns for Airports, Shopping Centers, and Industrial Plants (Project 7-1) 116 pp. $5.20
25	Potential Uses of Sonic and Ultrasonic Devices in Highway Construction (Project 10-7) 48 pp. $2.00
26	Development of Uniform Procedures for Establishing Construction Equipment Rental Rates (Project 13-1) 33 pp. $1.60
27	Physical Factors Influencing Resistance of Concrete to Deicing Agents (Project 6-5) 41 pp. $2.00
28	Surveillance Methods and Ways and Means of Communicating with Drivers (Project 3-2) 66 pp. $2.60

* Highway Research Board Special Report 80.

THE NATIONAL ACADEMY OF SCIENCES is a private, honorary organization of more than 700 scientists and engineers elected on the basis of outstanding contributions to knowledge. Established by a Congressional Act of Incorporation signed by President Abraham Lincoln on March 3, 1863, and supported by private and public funds, the Academy works to further science and its use for the general welfare by bringing together the most qualified individuals to deal with scientific and technological problems of broad significance.

Under the terms of its Congressional charter, the Academy is also called upon to act as an official—yet independent—adviser to the Federal Government in any matter of science and technology. This provision accounts for the close ties that have always existed between the Academy and the Government, although the Academy is not a governmental agency and its activities are not limited to those on behalf of the Government.

THE NATIONAL ACADEMY OF ENGINEERING was established on December 5, 1964. On that date the Council of the National Academy of Sciences, under the authority of its Act of Incorporation, adopted Articles of Organization bringing the National Academy of Engineering into being, independent and autonomous in its organization and the election of its members, and closely coordinated with the National Academy of Sciences in its advisory activities. The two Academies join in the furtherance of science and engineering and share the responsibility of advising the Federal Government, upon request, on any subject of science or technology.

THE NATIONAL RESEARCH COUNCIL was organized as an agency of the National Academy of Sciences in 1916, at the request of President Wilson, to enable the broad community of U. S. scientists and engineers to associate their efforts with the limited membership of the Academy in service to science and the nation. Its members, who receive their appointments from the President of the National Academy of Sciences, are drawn from academic, industrial and government organizations throughout the country. The National Research Council serves both Academies in the discharge of their responsibilities.

Supported by private and public contributions, grants, and contracts, and voluntary contributions of time and effort by several thousand of the nation's leading scientists and engineers, the Academies and their Research Council thus work to serve the national interest, to foster the sound development of science and engineering, and to promote their effective application for the benefit of society.

THE DIVISION OF ENGINEERING is one of the eight major Divisions into which the National Research Council is organized for the conduct of its work. Its membership includes representatives of the nation's leading technical societies as well as a number of members-at-large. Its Chairman is appointed by the Council of the Academy of Sciences upon nomination by the Council of the Academy of Engineering.

THE HIGHWAY RESEARCH BOARD, organized November 11, 1920, as an agency of the Division of Engineering, is a cooperative organization of the highway technologists of America operating under the auspices of the National Research Council and with the support of the several highway departments, the Bureau of Public Roads, and many other organizations interested in the development of highway transportation. The purposes of the Board are to encourage research and to provide a national clearinghouse and correlation service for research activities and information on highway administration and technology.