Getting the Most Out of a Freeway System

A. C. Estep and Karl Moskowitz*
California Department of Transportation

Responsibility for the operations, operational analysis, and planning of operational improvements of the California state highway system lies with the Department of Transportation. These three activities are defined, and their application in a program to upgrade and control the Los Angeles freeway network is discussed. Methods of controlling freeway traffic (ramp metering, freeway widening, new construction) are described.

California is fortunate enough to have constructed a major portion of a freeway system that provides a higher level of transportation service for more people than has ever been provided anywhere. Nearly 40 percent of all ground transportation in California now occurs on 3,800 miles of completed freeways. We owe this to an exceptionally farseeing legislature that not only established the freeway and expressway system in 1959 but followed through with a resolution calling for a study of how this system would be operated. This study (1) was based on the premise that "the state has the continuing responsibility to assure that the use of highway plant shall be efficient, safe, convenient."

Among the conclusions reached in this study were the following:

1. Traffic operations covers those things that are done after a highway is built that enable motorists and pedestrians to move safely and with a minimum of delay. All departments of the transportation agency contribute to the fulfillment of this function.
2. In the years ahead, the highway operations function will become increasingly important, and it will consume a larger share of the total funds available for highway purposes.
3. The so-called freeway problem is actually an aggregate of several problems. Some occur only on certain sections of the system; others are more or less general. Among them are recurrent congestion, congestion due to special events or accidents, accident prevention, disabled vehicles, directional signing, and law enforcement.
4. The transportation agency should consider the use of helicopter patrol of urban freeways, particularly during peak periods of travel.
5. The Division of Highways should consider the installation of a communications cable as a part of the freeway system.
6. Plans should be developed for emergency closing of freeway entrances and exits by remote-controlled electronic devices.

*Deceased.

7. The California Highway Patrol and the Division of Highways should undertake studies and experimentation to determine methods for faster clear-up of congestion on the freeways.

8. A freeway operations task force should be established under the general supervision of the highway transportation agency with representation from the Division of Highways, Highway Patrol, and Department of Motor Vehicles.

9. Freeway traffic control is concerned with traffic control decision making. It covers the entire traffic control process from data collection to communication of control decisions to motorists.

From that legislative mandate, the present policy of the California Department of Transportation has evolved.

OPERATIONAL RESPONSIBILITY

Providing safer, more expeditious travel is the ultimate goal of both highway construction and operation. Although new highway construction (including widening projects) is the major means available to alleviate traffic congestion and accidents, operations and operational improvements can, where appropriate, accomplish the same goal at relatively low cost.

It is the policy of the California Department of Transportation to make maximum effective use of the state highway system through a program of traffic operations and operational improvements. The California Highway Patrol has responsibility for controlling traffic, communicating with motorists, and coping with emergencies; generally these activities are handled by mobile forces. To the California DOT, which is fundamentally responsible for the physical aspects of the highway, operations, operational analysis, and planning operational improvements are separate but interrelated activities. All are necessary to operate the highway system effectively.

For the purposes of this paper, which is concerned with facilities as distinguished from enforcement, these three activities are defined below.

Operations

Operations does not mean the construction of operational devices nor modification of the geometry of a highway. Rather, operations includes surveillance and control of traffic and response to changing traffic conditions in real time.

Instructions to drivers (such as the color of a traffic signal) and information for drivers (such as a radio message) are changed by the moment, based on what is happening at a particular time. (Conventional highway signs also instruct or communicate with drivers, but the real-time factor is absent.)

The decision to install a traffic signal and the installation itself are planning, design, and construction in the traditional sense. The timing of the traffic signal, i.e., adjusting the dials, is operations. For another example, preliminary engineering and constructing a freeway ramp control project are not operations, but maintaining continuous surveillance of traffic on the affected freeway and adjusting ramp-metering rates are.

Operational Analysis

Operational analysis is the study of highway traffic. It can include things such as an inventory of freeway congestion, analysis of what happens to one bottleneck when another one is unplugged, and evaluation (ahead of time) of alternative methods of alleviating bottlenecks. Mathematical relationships that take into consideration items such as rate of flow, cumulative storage of vehicles by location in the system, highway capacity, density, and travel time come under this heading. This kind of analysis often becomes preliminary engineering for an operational improvement. It is also necessary input for real-time traffic control such as traffic signal settings. It is a kind of engineering somewhat different from conventional civil engineering, which has comprised the bulk of our work in the past.

Operational Improvements

The program for operational improvements involves specific projects designed to add to or better the existing system. It includes

1. Installing ramp-metering devices,
2. Placing changeable message signs,
3. Adding auxiliary lanes,
4. Providing left-turn storage lanes,
5. Restriping highways to provide added lanes,
6. Placing channelization devices,
7. Widening bottleneck locations, and
8. Installing traffic signals.

Although operational improvements were routinely accomplished when deficiencies became obvious, the process of identifying the possibilities and proposing improvements can be more organized and systematic. Responsibility for conducting operational activities and developing operational improvements must be clearly defined. Although the specific organizational structure for accomplishing this work may vary among transportation departments, or even districts, it is essential that the function be identified and adequately staffed. Only in this way will worthwhile improvements not be overlooked and will proper priority be established for each project in order to maximize the payoff in this area.

PROGRAM TO UPGRADE AND CONTROL THE LOS ANGELES FREEWAY NETWORK

In accordance with the concepts defined, a freeway operation unit was organized in Los Angeles in 1965. During the first year, the two-man staff mainly defined the scope of work and made recommendations for an action plan and proposed staffing, scheduling, and costs. The action plan called for operational projects that included a freeway congestion inventory to determine

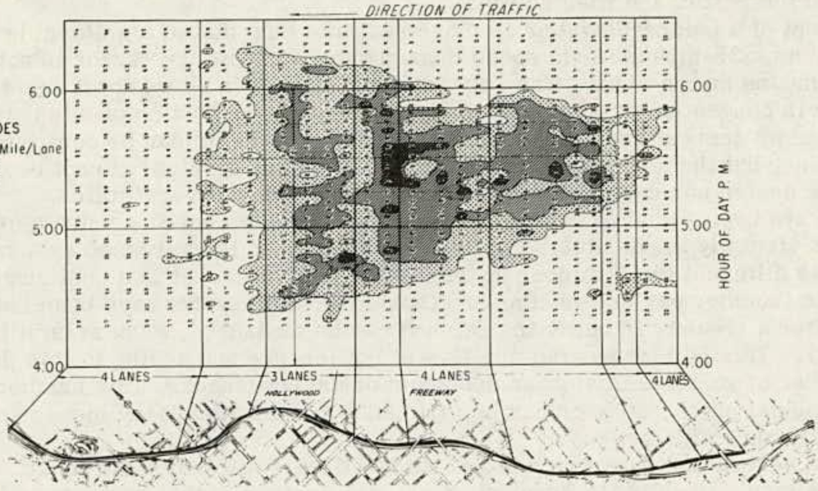
1. Location of problem spots,
2. Duration and extent of problems,
3. Quantitative estimates of travel time and speeds on various sections of the system,
4. Estimates of traffic demand at bottlenecks (where existing counts do not represent demand), and
5. The relationship of one bottleneck to another;

an analysis of specific bottlenecks (using inventory as base data); and a correlation of design and planning functions. The action plan also called for the following planning and research projects:

1. Evaluation of methods of ramp control,
2. Survey of nonrecurrent congestion, and
3. Communication with the motorist (signs, radio, etc.).

The first analysis of specific bottlenecks was done in 1966, concurrently with the development of scope of the study, and resulted in a manually operated metering signal at one ramp and peak-period closure (also manned) of another ramp on the Hollywood Freeway. This project was followed, in 1968, by a system of five metered ramps and two ramp closures on the Harbor Freeway. These initial projects were very successful

Figure 1. Typical density chart.



in reducing overall delay on the freeway and in the involved corridors as a whole.

In 1969 the weekday congestion inventory was completed, and it was determined that 170 directional miles of freeway (out of approximately 700 directional miles at that stage of development) were subject to recurrent congestion during peak periods. This inventory was primarily based on peak-period aerial photography, using 35-mm hand-held cameras. The results are summarized on density contour maps showing time of day on the Y-axis, location along the freeway on the X-axis, and isodensity contours on the Z-axis. Forty-seven separate areas of congestion containing 106 individual bottlenecks and resulting in 1,000,000 vehicle-minutes of daily delay were identified and quantified in this initial study (1967-69 data). (A sample density chart is shown in Figure 1.)

The individual inventory reports on these bottlenecks were the basic data for a report issued in July 1970 (2), which concluded that "current manifest demand" could be accommodated by eliminating those bottlenecks at a cost of $115 million. The report was updated in 1972, and, like everything else, the cost went up. In 1972 the estimated cost was $160 million. We are now estimating $172 million, but it will probably exceed $200 million by the time we are through. This must be viewed in comparison with the $2.5 billion that was originally invested in the freeway system. In other words, we found that we could make the $2.5 billion system work for an additional investment of about 8 percent. In this context, work means essentially free-flowing traffic for 24 hours a day instead of 21 or 22 hours a day.

The ramp control and interim widening stages of this program are scheduled to be complete in 1977 or 1978. Preferential treatment for buses is an integral part of the program. In cooperation with the Regional Transit District, bypass lanes are provided at appropriate controlled entrance ramps, so that the advantages of exclusive bus lanes are obtained without reducing the capacity of the freeway. In fact, bus travel on the controlled sections is actually faster than it would be in a designated reserved lane on an uncontrolled freeway, and the thorny problem of access to the reserved lane is eliminated.

BALANCED FREEWAY SYSTEMS

It is promising that we may be able to construct and operate a truly balanced system of freeways. In this sense, balance means the relation between demand and capacity so that operating conditions, or quality of flow, on the freeways will be more uniform geographically and by time of day. Balance also includes balancing the load between freeways and the rest of the road network.

The concept of a balanced system is more realistic than that of a uniform level of service such as a 35-mph uniform speed throughout a system. The latter is not achieveable by any means known to us. The only reason traffic ever slows down below 45 mph is that there is congestion or stop-and-go traffic ahead. Congestion of this nature is always caused by demand exceeding capacity. The arrival rate can be controlled by ramp metering, but the demand-capacity ratio, or level of service, cannot be controlled by geometric design nor can it be uniform along a stretch of several miles.

It has always been a goal to balance demand and capacity, but an accurate forecast of demand is virtually impossible. [In this paper, demand is expressed as a rate of flow. This is different from number of trips per day or hour. It is a continuously changing rate (number per unit of time).] Therefore, bottlenecks have been inevitable. However, after a freeway is built, the demand can be measured, even when it is greater than capacity. This is done by counting flow at bottlenecks and adding to this the time-related number of vehicles in storage upstream of the bottleneck. This number can be obtained by aerial photography or can be estimated mathematically by measuring speeds and/or density along the freeway.

There are several reasons why demand rate of flow is difficult to forecast. First, of course, is man's inability to forecast anything. Another is that land use and travel patterns almost always are radically altered after the forecast is made by transportation planners. Furthermore, even if perfect forecasts could be made, rate of flow is by definition a quotient of number (of vehicles) divided by the time used to service this number. It is this latter concept that this discussion addresses.

At this point it may be desirable to define some terms. Manifest is a term that can be applied to demand, congestion, and delay. Counting traffic on the road or in the network, photographing congestion and counting stored cars, and measuring speeds for the purpose of describing quality of flow are all measurements of manifest phenomena—things we can see or get our hands on.

Latent can be defined most simply by example. If the service rate in the morning peak period is improved, people simply leave home later. The demand was always there, but it was invisible. There is a latent demand for trips on many urban freeways to be compressed within a shorter time period although the number of trips remains the same. This increases the required rate of flow and creates manifest congestion. If the latent demand is dampened or held constant, the congestion will not be manifest or will be manifest at a different location. When the time spread of trip demand approaches zero, the capacity necessary to absorb the flow rate will approach infinity.

Similarly, if one route becomes more attractive (has less delay) than another, the demand for the attractive route will rise until the capacity of the less attractive route is reached and delay occurs on the attractive route. Conceptually it is possible to keep adding capacity to one route or system of routes, thus making it more and more attractive, until there is no traffic (or almost no traffic) on the less attractive route or system. It is not economically possible to provide enough capacity on one route to dry up all competitive routes. Freeway entrance ramp control offers a solution to this dilemma.

FREEWAY CONTROL

Ramp control or metering consists of traffic signals at entrance ramps that control freeway input at a rate the freeway is capable of handling. It does not increase the capacity of the freeway, but, by diverting or storing traffic, it allows more traffic to use the freeway upstream of that point. Thus, the throughput of the freeway or the rate of accommodating vehicle-miles of travel is increased. Use of electronic sensors and computers makes the signals more responsive to fluctuations in traffic flow. Because the rate of flow on the freeway is less than capacity at all points along the controlled section, the level of service or travel speed is relatively high. Speed seldom drops below 40 mph and, if it does, only for short stretches.

Ramp Control on Existing Freeways

A congestion inventory of existing freeways should be made and kept current. From this inventory, an estimate can be made of the potential increase in throughput (vehicle-miles per unit of time) that can be achieved by controlling the flow rates on entrance ramps. An estimate is also made of the effect on surface streets and other routes that may be used as alternates by traffic diverted at the ramp entrances. In many instances, the majority of traffic diverted at controlled ramps has actually been accommodated on the main line of the freeway being controlled. In other words, many of the same trips use the freeway, but enter it at some upstream point, thus getting a longer and much faster freeway ride. This is accomplished by increased throughput. In all cases, the actual number of vehicles using a controlled ramp during a peak period decreased. (We say this to dispel the fear that ramp queues will be intolerably long.)

Estimates are also made of the savings in travel time that will be experienced by present users of the freeway and new users, and these are compared with the detriment in travel time suffered by the traffic stored at the ramp signals or diverted to other routes. If the net savings is large enough, the metering system should be installed (3).

The congestion inventory will also reveal certain imbalances in demand-capacity ratios that can be corrected by geometric improvements (generally by widening or adding auxiliary lanes in bottlenecks).

Metering systems should be installed on some freeway sections that are not currently suffering peak-period congestion but that prove through continuous surveillance to be approaching this state. The reason for this is to obviate the shock in travel patterns that occurs when ramp metering is installed on a congested freeway.

The magnitude and suddenness of changes in trip patterns that can happen were illustrated when on-ramp volumes on 5 miles of the Harbor Freeway were reduced by 1,400 vph the day the ramp signals were turned on. Although the travelers readily adapted to the changes and were pleased with the results, a gradual imposition of control would have caused less disruption.

New freeways should be planned so that surveillance and ramp control can be implemented with minimal revision. In certain cases, the hardware can be installed as part of the major construction contract. If metering is inaugurated before congestion sets in, the advantages that are always observed when a freeway is first opened to traffic will be preserved for an indefinitely long period, with little noticeable deterioration to the metered traffic.

As noted earlier, geometric bottlenecks are inevitable because of unforeseen changes in demand, among other reasons. Input can be controlled so that congestion does not occur even when geometric imbalance exists. However, this is undesirable because it results in underutilization of a large portion of the freeway. The desirable thing to do is to measure the actual demand at all locations along the freeway and to keep this measurement up to date. When geometric bottlenecks are identified they should be corrected.

Freeway Widening

When a freeway is widened to accommodate existing congestion, ramp control should be installed. This is the only way to ensure that latent demand will not congest the newly widened freeway. The ramp-metering plan can be very liberal to begin with, resulting in short or nonexistent ramp queues and delay. Geometric modifications of the ramps and electrical conduits should be an integral part of the widening contract. The only problem with this procedure is that the user benefits of the ramp control cannot be evaluated independently of the benefits of widening.

FREEWAY SURVEILLANCE BY COMPUTERS

There is nothing in the California experience to indicate that computer surveillance of traffic flow will drastically affect highway safety. Computer surveillance can assist in early detection and management of traffic accidents. The earlier an accident or incident is removed from the roadway, the less chance there is for a second accident to occur as a result of congestion created by the first one.

But nothing will ever take the place of efficient incident management, and this can only be accomplished by a management team, the key member of which is the law enforcement officer or highway patrolman on the scene.

The entire California freeway system of some 3,800 miles is under routine surveillance by the California Highway Patrol. On 42 miles of this system, the patrol is assisted by electronic surveillance. The computer is only one element of electronic surveillance, albeit a key element.

The Los Angeles Area Freeway Surveillance and Control Project—the 42-mile loop—is a large-scale experiment to determine what can be done to reduce delay, reduce accidents, and relieve motorist frustration caused by nonrecurrent congestion. It is both an operating system and an experimental system. Planning for this experimental system started in 1968, and it became operational in late 1971. We have now had 3 years of operating experience.

To a large extent, the project was a response to mounting public concern that everything possible be done to maximize the operational effectiveness of the existing highway plant through the use of state-of-the-art electronic technology. Forty-two miles of the heaviest traveled urban freeways in California were selected as the site of the project in order to make the test valid and to make it large enough in scope to be realistic. It should be noted, however, that 42 miles is only 10 percent of the Los Angeles urban freeway network, and in this respect the experiment was conservative in scope.

All the 42 miles are eight-lane or 10-lane freeways, and about 700,000 trips per day use one part or another. The highest traffic volume in this system reaches 240,000 vpd, although the average for the whole loop is about 120,000 to 150,000 vpd.

Although operations engineers have always kept track of what is happening on freeways (i.e., traffic volumes, accidents, and congestion problems), this project differs in that we get information by the moment, instead of by the day or the year, and all over instead of at random locations. The project consisted of four major elements.

Continuous data are collected for operations research. We hoped to obtain much more accurate estimates of congestion and of the daily variation in flow and congestion than had previously been available. Actually, data processing became so complex that we have been unable to produce meaningful summary data that could be used by researchers or management.

Incidents that have caused a difference in flow or congestion are detected early. It was our goal to detect 90 percent of the incidents within 5 minutes, with an average of 1.5 minutes. We came close to that goal, at the expense of some false alarms. Incidents include accidents and events such as stalled vehicles and gravel spilling from a truck, as well as major problems such as a truck tipping over or a landslide.

Early detection is not so important as management of incidents. Electronic surveillance per se does nothing to improve incident management. Communication between the data surveillance center and the field command must be greatly improved if the surveillance is going to be meaningful.

Information for the motorist is transmitted via commercial radio. However, the changeable message signs are used so infrequently and it is so difficult to make them timely, meaningful, and accurate, that we are not planning to expand their use. Even when they are timely and accurate there is usually nothing the driver can do that he could not do in the absence of such signs.

We are able to make ramp control more responsive to random changes in flow. This feature of the experimental surveillance system is being continued on an operational basis.

Surveillance, especially electronic surveillance, means keeping track of what is happening in real time, and in the Los Angeles project it was addressed primarily to unusual events or nonrecurrent congestion, as opposed to recurrent congestion.

Electronic surveillance of a freeway system does not affect traffic flow except very indirectly. The only way to control traffic on a freeway is by controlling it before it gets on with ramp signals. As opposed to surveillance, ramp control has a positive and dramatic effect on the flow of traffic on the freeway, and everybody knows that something has been done. Real-time surveillance is not necessary in order to run a ramp control system. It is necessary, however, to keep track of operating conditions and to adjust ramp-metering rates from time to time. For this reason, it might be desirable to install electronic surveillance even if it is used only to furnish data for off-line adjustments of the metering plan.

CURRENT AND FUTURE RESEARCH

Some of the problems that we are working on can be solved by research and some are engineering design problems that can only be solved by experience (or experiment). Some may never be solved.

We are now developing our third generation of ramp controllers. The first was a pretimed controller run by clockwork and had a maximum of three metering rates. This controller could not read demand or passage of a vehicle and therefore needed a yellow interval, for it could turn green when a car was still a long way upstream and then turn red before the car arrived.

For the second generation, the controller has three adjustable metering rates and can read whether a detector upstream or downstream of the ramp signal is occupied. Thus we can use red-green sequence with no yellow and get more definite one-by-one metering. The adjustable metering rates can be preset by time of day or can be called up by remote supervisory computer control.

The signal goes from steady green (in the off-peak mode) to a 3- or 4-second yellow when the metering mode starts. From then until the metering period is over, the signal rests on red so that approaching cars see a red light if there is no queue, and the yellow is unnecessary. The signal turns green when (a) there is a car waiting (on the calling detector) and (b) sufficient time has elapsed since the last green so that the allowable metering rate is not exceeded. The light stays green until the car to be served crosses the canceling (passage) detector. This sequence of decisions allows for variable reaction times among drivers and at the same time the second car in line cannot start while the light is green (because the first car has not moved). The second generation controller costs about $1,200.

It should be noted that the second generation controller requires external equipment and telemetry to change metering rates in real time (although it can operate independently if metering rates are set by clock time). The third generation, which is being specified for projects now under design, is a microprocessor that costs $1,500 to $2,000 and can do everything the second generation does plus the following:

1. It can adjust its own metering rate in response to local traffic parameters;
2. It can batch data for transmission to a central computer (thus making telemetry requirements less troublesome); and
3. It can check out malfunctions in local hardware.

One of our earliest projects (Chula Vista) used a homemade analog processor that had continuously variable metering rates responsive to occupancy on the freeway upstream of the ramp.

Of the hundred or so ramps under control in California, about 80 percent have preset metering rates. The only ramps that have metering rates responsive to main-line fluctuations in flow, on a system basis, are the unique main-line meter on the Bay Bridge and two systems on the 42-mile loop that are supervised by a large computer (Sigma 5) connected by leased telephone lines. We feel that interconnection and traffic responsiveness should be beneficial and that there are more efficient methods than using large computers with large data transmission requirements. But when we start designing alternative interconnect plans, they turn out to be so expensive that there is considerable doubt about their cost effectiveness.

One of the reasons for systemwide control with surveillance (feedback of traffic parameters) is to save labor. That is to say, the manpower requirements for manual off-line adjustments of 500 to 1,000 metering plans are formidable. But now we are beginning to wonder whether electronic surveillance will actually enable us to reduce total manpower requirements. First, it takes considerable manpower to operate and maintain the surveillance hardware. And this particular type of manpower is very difficult to train and keep. Second, we do not know exactly how to read the traffic parameters that electronics can measure, nor what algorithms to use to change or update control strategies. Third, a surveillance system that would tell us what is going on on the surface street portion of the corridor would be so complex and costly that we would never get it built.

We need theoretical or analytical models to tell us what to do to optimize flow in the corridor.

Main-line metering, or metering freeway-to-freeway connectors, has barely been scratched.

In conclusion, I would say that we are making progress but there are enough problems remaining to keep research teams as well as operations engineers busy for many years.

REFERENCES

1. Services of the State Government for the Motoring Public. 1963.
2. A Program to Upgrade and Control the Los Angeles Freeway Network. California DOT, July 1970.
3. Freeway Ramp Control—What It Can and Cannot Do. Traffic Engineering, July 1969.