ENTRANCE RAMP CONTROL FOR TRAVEL-RATE MAXIMIZATION IN EXPRESSWAYS* 

CHENG-I CHEN†
Autonetics Division, Rockwell International, Anaheim, California 92803, U.S.A.

J. B. CRUZ, JR.
Coordinated Science Laboratory, University of Illinois, Urbana, Illinois 61801, U.S.A.

and

J. G. PAQUIET
Department of Electrical Engineering, Laval University, Quebec, Canada

(Received 20 October 1972; in revised form 14 January 1973)

Abstract—The steady state optimization of expressway system operations when peak demands exceed capacities is investigated. An alternative approach of using travel-rate or average traffic flow of the whole expressway system instead of input rate as the criterion of optimization is adopted. This leads to a linear programming problem. Residual effects of peak-period traffic are taken into consideration in computing total time to return to normal traffic pattern. For the example considered, this total clearing time for a control scheme using travel-rate maximization is slightly less than that for one using input rate maximization. In addition, travel-rate maximization control has a tendency to discourage vehicles making short trips to use the freeway system during the peak period.

1. INTRODUCTION

Several different models have been proposed to describe the relationship between space mean speed $u$ and density $k$ for a given section of a directional expressway (Greenshields, 1934; Lighthill and Whitham, 1955; Edie, 1961; Greenberg, 1959; Herman and Rethery, 1962; Pipes, 1965). Using Greenshields model one obtains

$$u = u_t \left(1 - \frac{k}{k_s}\right)$$  \hspace{1cm} (1)

where $0 \leq k \leq k_s$; $u_t$ and $k_s$ are positive constants, known as the freepoint and the jam density, respectively. Observations have shown that this model closely represents real traffic situations over a wide range of densities (Greenshields, 1934; Edie et al., 1963; Kreer and Goodnuff, 1967). At any point of the freeway the flow of traffic $q$ is given by

$$q = uk.$$  \hspace{1cm} (2)

Thus

$$q = u_t \left(k - \frac{k_j}{k_s}\right).$$  \hspace{1cm} (3)

Clearly $q$ reaches its maximum $q_{max}$ at $k = \frac{k_j}{k_s}$ and the corresponding traffic flow is $\frac{1}{2}u_t k_j$. The quantity $q_{max}$ is called the capacity of that section of the expressway. As $u_t$ and $k_{j}$ may be different from section to section, the capacity of the expressway will also be different for different sections.

Due to limitations on the amount of traffic the system can handle, traffic density may exceed its critical value $\frac{k_j}{k_s}$ during peak periods if no control is applied. This build up will in turn reduce the traffic flow further. Thus, it is desirable to control the injection rates at entrance ramps so as to keep the densities of the traffic along the system below critical densities.

One method of achieving this traffic control is to consider it as a two-level optimization problem with the steady-state (long-term) optimization as the first level objective and the dynamic (short-term) regulation around the optimal steady-state solution as the second level control (Yuan and Kreer, 1969; Isaksen and Payne, 1973). In this paper only the first level optimization problem will be considered.
In Wattleworth and Perry (1965) and Wattleworth (1967), it has been shown that one possible way of implementing first-ave. control is to choose injection rate as the criterion of optimization. Some relations between travel time and injection rate have been discussed previously (Wattleworth and Perry, 1965). Only fixed-final-time problems have been considered, and effects of peak period control on post-peak period traffic have not been taken into account. Various variations of injection rates have been chosen by others as a criterion of optimization. In Yuan and Kretz (1969), length of queues and average demand have been used as weighting functions.

In this paper, average flow as the criterion of optimization will be studied. This choice is based on the observation that utilizing the system as near as possible to its full capacity will allow traffic to move more efficiently.

2. TRAVEL-RATE MAXIMIZATION

Expressways with multiple entrances and multiple exits will be considered. A section of expressway is defined as the part of the expressway between any two adjoining entrances and exits.

The Greenshields model of space-mean speed versus density is assumed to hold. The amount of flow the system can handle is thus bounded. If the capacity is not uniform along a section of expressway due to changes in numbers of lanes or other geographical configurations, the minimum of the maximum possible flow in that section is taken as the capacity of that section. In other words, the capacity of a section of expressway is the amount of traffic that can be handled at its bottleneck without causing congestion. Any vehicle which intends to travel through the system constitutes demand to the system. The total demand at a certain point in the expressway is the number of vehicles which intend to use that specific portion of the expressway at a certain instant of time. If demands are smaller than capacities at all sections of the system during the entire interval of consideration, no traffic congestion can occur even with no control. However, when the demand at certain sections exceeds the capacity there, high density of traffic will result in the expressway. This high density of traffic may further decrease flow and as a result cause congestion in upstream sections, resulting in a breakdown of the system. To avoid such a congestion and possible total breakdown of the system, it is desirable to control the injection rates of the vehicles at different ramps. The steady-state optimization of such traffic control problems when incoming demands exceed the capacities of the expressway for at least one section of the system will be studied.

Entrances in the system under consideration will be labeled as \(i_1, i_2, \ldots, \) etc., and exits labeled as \(e_1, e_2, \ldots, \) etc. Consider a vehicle which enters the system at a certain entrance \(i_j\) located at \(x\) at time \(t_0\). As this vehicle travels along the system, it poses a demand along the system. This demand reaches point \(z\) at time \(t = t_0 + \tau\) where \(\tau\) is the time required for it to travel from \(x\) to \(z\). In general, \(\tau\) is a function of densities along the trajectory. Thus to determine the demand at \(z, \tau\), the past history of the densities, injections, and the relationship between densities and velocities along the expressway are needed. Only the average behavior is considered and possible deviations of individual behavior are considered as disturbance. Such disturbances are to be considered in the second-level control (dynamic regulation) problem only.

When the system reaches steady state, the density of the traffic at each section is constant with respect to time. The traffic flow is also time-invariant since it is uniquely determined by the density.

In order to reduce traveling delay, it is highly desirable to utilize the limited expressway capacity as close as possible to its full capacity. With this consideration, it is desirable to maximize the average flow along the expressway. The flow at each point of the expressway is \(q(t, z)\). The average flow or travel-rate in the system is equal to

\[
J(t) = \int q(t, z)dz
\]

(4)

Since in the steady-state operation \(q\) is constant over each section of the expressway, (4) reduces to

\[
J = \frac{1}{D} \int qdz
\]

(5)

where \(d_i\) is the length of the \(i\)th section and \(D\) is the total length of the freeway. Clearly, maximizing the travel rate is equivalent to maximizing the total travel rate \(JD\) since \(D\) is a constant.

Let \(a_{ij}\) be the portion of the traffic which enters the system at entrance \(j\) and leaves at exit \(i\). In general, one obtains

\[
e_i = \sum_{j=1}^{r} a_{ij} u_j \quad i = 1, \ldots, r
\]

(6)

where \(u_j\) is the injection rate at entrance \(j\), \(e_i\) is the departing rate at exit \(i\). If exit \(i\) precedes entrance \(j\), \(a_{ij} = 0\). Also, \(\sum_j a_{ij} = 1\) for all \(i\). The \(r \times r\) matrix \(A\) where \(i\) is the number of exits and \(r\) is the number of entrances, whose elements are \(a_{ij}\), is called the origin-destination matrix, and it is assumed to be time-invariant.

The origin occupation equation is

\[
e_i = \sum_{j=1}^{r} b_{ij} u_j
\]

(7)
where $h_i$ is the portion of vehicles which enter the system at entrance $i$ and utilize route $i$. The elements $h_{ij}$ form an $s \times r$ matrix $B$ where $s = 1 + r - 1$ is the number of sections. It is related to matrix $A$ by

$$h_{ij} = \sum_i a_{ij}$$

where section $i$ precedes exits $j$'s.

Substituting (7) into (5),

$$F = \frac{1}{D} \sum_i d_i \sum_j h_j u_j = \frac{1}{D} \sum_i \left( \sum_j h_{ij} d_i \right) u_j = \frac{1}{D} \sum_i \left( \sum_j h_{ij} d_i \right) u_i. \quad (8)$$

The problem of finding an injection rate for each entrance to achieve travel-rate maximization is then a mathematical programming problem of determining $u_i$ such that (8) is maximized subject to the constraints equation (7) be satisfied

$$0 \leq u_i \leq (a_{ij})_{\max} \quad \forall i, \quad i = 1, 2, \ldots, n \quad (9a)$$

$$0 \leq k_i \leq \frac{1}{(k_i)_{\max}} \quad \forall i, \quad i = 1, 2, \ldots, s \quad (9b)$$

$$q = u_i \left[ \frac{k_i}{(k_i)_{\max}} \right] \quad \forall i, \quad i = 1, 2, \ldots, s \quad (9c)$$

where $(a_{ij})$ is the free speed of section $i$, $(k_i)_{\max}$ is the jam density of section $i$, and $(a_{ij})_{\max}$ is equal to demand at entrance ramp $i$ if there is no queue there and is equal to the maximum possible injection rate of entrance ramp $i$ if there is a queue.

Equation (9d) can be rewritten as

$$k_i = \frac{k_i}{\sqrt{k_i^2 - 4q_i u_i}} \quad (9d)$$

The second inequality of (9c) can be satisfied only when the negative sign is chosen in (9d). Also since $k_i$ must be real and positive, it follows that

$$0 \leq q_i \leq (k_i)_{\max} \quad (10)$$

Thus equation (9) becomes

$$0 \leq \sum_i h_{ij} u_j \leq 4(u_i)(k_i) \quad \forall j, \quad j = 1, 2, \ldots, s \quad (11a)$$

$$0 \leq u_i \leq (a_{ij})_{\max} \quad \forall i, \quad i = 1, 2, \ldots, n \quad (11b)$$

$$0 \leq k_i \leq \frac{1}{(k_i)_{\max}} \quad \forall i, \quad i = 1, 2, \ldots, s \quad (11c)$$

$$k_i = \frac{k_i + \sqrt{k_i^2 - 4q_i u_i}}{2} \quad \forall i, \quad i = 1, 2, \ldots, s \quad (11d)$$

Since (11c) is satisfied when (11a) and (11d) are satisfied, the constraints which will affect the cost function $J$ are (11a) and (11b) only. Since the objective function $J$ is linear in $u$ and since the constraints (11a) and (11b) are linear this is a linear programming problem whose solution can be obtained by applying the simplex method or other widely used methods. Once $u$ is obtained, the density $k_i$ given by (11d) automatically satisfies (11c).

### 3. Clearing Time and Total Road Occupation

The period during which the incoming traffic demand is greater than the capacity of the expressway is called peak-period. The next period in which the incoming traffic demand returns to normal level but the traffic conditions is still affected by the residual queues is called post-peak period. Only after all queues disperse and the density of the traffic along the expressway returns to its nominal density does the traffic return to its normal pattern. It is assumed that normal traffic demand is less than the capacity of the expressway. The time interval between the initiation of the peak period and the termination of the post-peak period is defined as the clearing time $T$. It is the sum of the peak period and the post-peak period. The peak period is dictated by the specified demand and it is not affected by the type of control used. However, the post-peak period depends on the control. In the next section clearing times will be compared for the two cases when travel-rate and input-rate are used as optimization criteria. In both cases controls are applied only during the clearing time duration. The optimal control for the post-peak period may be different from that for the peak period, in each of the two cases.

When all queues are cleared and when the traffic densities have their normal values, the total number of vehicles in the freeway system is a constant. During the peak and post peak periods the total number of vehicles in the freeway and in the queues varies with time. If the total number of vehicles in the system is plotted with respect to time, the number of vehicles $V$ will increase during the peak period, decrease during the post-peak period, and return to the steady-state value at the end of the clearing time. Thus the clearing time is easily obtained if a plot of $V$ is available. The total number of vehicles $V$ in the system can be written as

$$V = \sum_i k_i d_i + \sum_i Y_i \quad (12)$$

where $Y_i$ is the queue length at entrance ramp $i$. Using (9d) and (11d), (12) can be rewritten as

$$V = \sum_i \frac{\left(k_i d_i \right)}{2} \left[ 1 - \sqrt{1 - \frac{4q_i u_i \left(k_i, \left(k_i \right)_{\max} \right)}{2}} \right] + \sum_i Y_i \quad (13)$$

The flow $q_i$ will remain constant during the peak period and it will be piece-wise constant during the post-peak
period. Transitions in $v_i$ could occur as queues are cleared. In general, $v_i$ increases during the peak period and decreases during the post-peak period.

Another index which we will use for comparing the two controls in the example of the next section is total road occupation defined as the area under the curve of $\int_{0}^{t} v(t) dt$ (14).

$$\text{Total road occupation} = \int_{0}^{t} v(t) dt$$

expressed in vehicle-hours. In the literature this is also known as actual travel time in vehicle-hours. After the end of the clearing time, total road occupation increases linearly with time.

4. EXAMPLE

Steady-state control of a simple hypothetical four-section freeway is studied. Figure 1 shows a schematic of sections of freeway for this example. There are three entrances, $i_1$, $i_2$ and $i_3$, and two exits, $e_1$ and $e_2$. Demands and origin-destination matrices for peak-period and normal conditions are given in Tables 1 and 2. Peak-period duration is assumed to be 30 min. It is assumed that the maximum possible injection rates at entrance ramps are large enough so that they are not active constraints. It is also assumed that controls are applied until the traffic returns to its normal condition.

Results for input-maximization and travel-rate maximization are given in Tables 3 and 4. The total queue length at the end of the peak period for the input-

| Table 1. Demands and origin-destination for peak period condition |
|-----------------|-----------------|-----------------|
| Entrance | Exit | $v_1$ | $v_2$ | Total demand |
| $i_1$ | $e_1$ | 400 | 3600 | 4000 |
| $i_2$ | $e_1$ | 200 | 800 | 1000 |
| $i_3$ | $e_1$ | 640 | 640 | 640 |

| Table 2. Demands and origin-destination for normal condition |
|-----------------|-----------------|-----------------|
| Entrance | Exit | $v_1$ | $v_2$ | Total demand |
| $i_1$ | $e_1$ | 320 | 2880 | 3200 |
| $i_2$ | $e_1$ | 120 | 480 | 600 |
| $i_3$ | $e_1$ | 600 | 600 | 600 |

| Table 3. Input rate maximization control |
|-----------------|-----------------|-----------------|
| | Peak-period | Post-period | Normal period |
| | 30 min | 45 min | period |
| Input rates | | | |
| Entrance ramp $i_1$ | 3400 | 3600 | 3200 |
| Entrance ramp $i_2$ | 1000 | 600 | 600 |
| Entrance ramp $i_3$ | 640 | 600 | 600 |
| Densities | | | |
| Section 1 | 77.4 | 100 | 60.7 |
| Section 2 | 100 | 78.7 | 83.4 |
| Section 3 | 81.5 | 73.6 | 60.4 |
| Section 4 | 100 | 80 | 65.4 |

Queue at the end of the peak-period entrance ramp: $v_1 = 300, v_2 = 200, v_3 = 100$.

| Table 4. Travel rate maximization control |
|-----------------|-----------------|-----------------|
| | Peak-period | Post-period | Normal period |
| | 30 min | 45 min | period |
| Input rates | | | |
| Entrance ramp $i_1$ | 3600 | 3600 | 3200 |
| Entrance ramp $i_2$ | 800 | 800 | 600 |
| Entrance ramp $i_3$ | 620 | 620 | 600 |
| Densities | | | |
| Section 1 | 100 | 100 | 60.7 |
| Section 2 | 100 | 100 | 63.4 |
| Section 3 | 82.68 | 82.68 | 68 |
| Section 4 | 100 | 100 | 65.4 |

Queue at the end of the peak-period entrance ramp: $v_1 = 200, v_2 = 100, v_3 = 10$. |
maximization problem is less than the corresponding values for the travel-rate maximization problem, as one would expect. However, the residual effect on post-peak period traffic for the former is greater than that for the latter. It takes 30 min for traffic to return to normal operating conditions for travel rate maximization while it takes 45 min for the other control scheme. This more time-consuming post-peak period traffic in the input maximization case is partly due to the presence of a queue on entrance ramp 1, which is longer than any other queue using travel rate maximization. Figure 2 shows a plot of \( F \) for the two cases. The rapid changes in \( F \) at the transition points are due to the differences in steady-state values of traffic densities for the normal, peak, and post-peak periods. The transition duration is approximately 100 sec which is the time it takes a vehicle to travel the first section. The transition time to change densities in the other three sections of the freeway is less than 100 sec. The mean free speed of section 1 is 72 m.p.h. Assuming an average of 100 v/ln (see Tables 3 and 4) the actual space mean speed in section 1 using equation (30) is 36 m.p.h. At this speed it takes 100 sec to go through the 1 mile section.

Figure 3 shows plots of the total road occupation for the two cases. As expected the road occupation during the peak period is less for the input rate maximization control. However, when the post-peak period is considered, total road occupation is less for the travel-rate maximization control for this example.

5. DISCUSSION

For an ideally balanced freeway system operation, the upper bound for \( k \), in (9c) should be attained for all index \( i \). Of course, this may not be completely achievable since the number of lanes is an integer and demands change from time to time. However, at the design stage, achieving balance of the system should be an important factor of consideration. As demands may change over a long period or upon completion of the system, the indices \( T \)'s of (9b), (9c) serve as indications as to where upper bounds for \( u_i \) and \( q_i \) are actually attained and where expansion of the existing system may be needed. These sections may not exactly coincide with sections where congestion would occur when the system is not under control. Also, since the origin-destination matrix may vary over a long period and it may be different during different seasons and different peak-periods of the same day, indices \( T \)'s of active constraints may also change.

In the paper it is assumed that the origin-destination matrix is time invariant. This assumption is generally not valid for an entire peak period although it is reasonable to assume similar patterns at the same hour of different weekdays. However, its time constant can be expected to be much longer than other time constants in the second-level dynamic regulation of the system. From a practical point of view as well as computational feasibility, they are considered as constant over a subperiod and the whole static optimization solution is adjusted from one subperiod to another subperiod. Recomputation of the optimum steady state is also necessary if certain queuing situations become unacceptable. As demands are assumed to be constant over each subperiod, accuracy is sacrificed if each subperiod is taken to be very long. On the other hand, if intervals of subperiods decrease, then in addition to computational and implementational difficulties, dynamic responses may play more important roles. Effects of dynamic responses are not taken into consideration in this paper.

As pointed out in previous studies, one of the potential drawbacks of a metering system is the build-up of vehicles at certain metered entrance ramps. One possible way of resolving this difficulty is to place an upper bound on the number of vehicles which are not allowed to enter the freeway (Wadsworth and Perry, 1965). In Wadsworth (1967) an additional constraint that queues be balanced is imposed. Still another alternative is to put additional penalties on queuing to the cost functions (Yuan and Kreer, 1969, 1971). A queuing situation will
always exist as long as the capacity of the freeway system cannot meet the demand.

The suggested control scheme in the paper is based on maximization of average traffic flow of the system when peak demand exceeds the capacity of the system. No attention has been placed on diverting any portion of the demand to streets or other portions of the system. However, as an important side effect of the control scheme proposed here, drivers who make short trips may be discouraged from using the freeway system during the peak period. Demands for short trips are weighted less than demands for long trips. Thus queues are more likely to develop in ramps with more short trips. Since expected waiting time at ramps is independent of the distance of travel, drivers of short trips tend to have larger waiting time-traveling time ratios. In the example, where there is queuing at entrance ramp $b$, many vehicles which travel from $a$ to $c$ may find it more expedient to use alternative arterial streets. In general, congestion on an arterial street does not decrease the flow rates on those streets as much as in freeway systems (Watts and Perty, 1965).

From the relationship between total road occupation of freeway flow (13) and (14), it is easy to see that a very slight decrease of flow has a very large effect on total road occupation when flow is very near the capacity of the freeway. Of course, a queuing situation will also be improved by discouraging short trips on the freeway during peak periods. On the other hand, input-maximization control tends to give priority to short trips since vehicles which make short trips occupy the system for less time and thus create more gaps to be reused downstream and increase the input rates.

It is also to be noted that in computing total road occupation, it is important to consider the residual effects in addition to the peak period. A control scheme which minimizes the road occupation for the peak period alone does not represent the real situation. In the previous example, even when no traffic is diverted, travel-rate maximization control yields a smaller total road occupation than input rate maximization control when the residual effects are taken into account. The residual effects depend on normal period demands as well. However, except for very simple cases, computation and comparisons of total road occupation are not easy because it may require different post-peak periods to clear up the remaining queues at different entrance ramps. Also, when the system under consideration consists of a great number of entrance and exit ramps, the roles of dynamic responses will be more significant in the computation of residual effects.

It would be interesting to find the optimal control for the case when total road occupation (14) is used as the objective function, and to compare the results for all three approaches. However, nonlinearity of equation (14) and limitations of current techniques for solving nonlinear programming problems suggest that the problem is extremely difficult.

6. CONCLUSION

The main philosophy of the maximum travel-rate control scheme is that traffic at entrance ramps be controlled in such a way that the system is handling traffic near its maximum capacity under certain known or predicted demand situations. This scheme has the property that once vehicles are admitted to the system they will be able to proceed most efficiently. This control scheme is most attractive when peak demand exceeds the capacity of the system by a large margin or when the duration of peak demand is long, which is the case in most peak-period traffic situations. For the example considered in the paper, it was found that although the peak-period road occupation is larger than that corresponding to the input rate maximization scheme, the total road occupation is less if the total clearing time is considered.

REFERENCES


