The Analysis and Design of Freeway Entrance Ramp Control Systems

H. NATHAN YAGODA, Computran Systems Corporation, Wayne, N.J.; and
LOUIS J. PIGNATARO, Polytechnic Institute of Brooklyn

This paper describes some of the results obtained from an analytic study undertaken as part of the Gulf Freeway Surveillance and Control Project in which the on-line, dynamic control of individual entrance ramps is investigated. For this study 2 alternate control philosophies are considered. The first philosophy concerns the control of ramps on which a string of one or more vehicles is released when a suitable merge opportunity arises and when all previously released vehicles have successfully merged into the freeway stream. The second control philosophy concerns ramps on which the requirement that the ramp be cleared of all previously released vehicles is relaxed to increase the merge capacity of that ramp. In this second case, the controller may release an additional vehicle whenever the expected delay associated with attempting a merge into a detected gap is less than the expected delay associated with waiting for the ramp to clear. Here the controller is a sequential decision-maker that evaluates the expected delay associated with all previously released vehicles that have not yet merged and reflects this information into the control process. Fixed-time and demand-capacity metering controllers are special cases of the control system analysis presented.

•SEVERAL FREEWAY CONTROL projects are presently under way in the United States, and each is attempting to improve the operational characteristics of a freeway by use of an entrance ramp control system. However, the operational modes for the respective control systems and the rationale that underlies individual control system designs appear to be quite different. As a result of this diversity, there appears at first to be no clear-cut design philosophy with which to approach a new access ramp control system design problem. Because of this deficiency, a study group at the Polytechnic Institute of Brooklyn undertook the development of a more unified design theory for dynamic control systems for freeway entrance ramps. The study, initiated in August 1967, was part of the Gulf Freeway Surveillance and Control Project conducted by Texas A&M University in Houston, Texas, for the U.S. Bureau of Public Roads. The portion of that study described in this paper concerns the design of a dynamic control algorithm for an arbitrary entrance ramp configuration, subject to the assumption that control is exercised by use of a green-amber-red traffic signal located at a fixed position on the ramp. Control systems designs that utilize other controller configurations, such as speed signs or multiple signal stations, are not included in this presentation.

For the type of systems considered, 2 identifiable primary functions are (a) to improve the merge service offered to vehicles that enter the freeway and (b) to improve the operation of the freeway. The first function is performed by any system that better coordinates the arrival of the merge vehicle at the merge zone with the availability of a high-quality merge opportunity, and the second function is performed by any system that improves the freeway operational characteristics such as volume, speed, density,

Paper sponsored by Committee on Freeway Operations and presented at the 49th Annual Meeting.

accident rate, and occupacy. These 2 functions are not always compatible, but there is evidence to indicate that compatible solutions do exist. In particular, observation of the Gulf Freeway reveals that operation under conditions of ramp control (6) yields the following:

1. The rush period volume increased by about 10 percent;
2. The speed on the test section increased by about 30 percent;
3. The average travel time over a 5-mile section decreased from 16 to 11 minutes; and
4. The number of accidents during the 2-hour morning peak decreased from 145 to 75 per year for the 3 inbound lanes over the 6.5-mile controlled section.

DYNAMIC RAMP CONTROL

To understand the nature of the control system design problem for a freeway entrance ramp required that initial consideration be focused on the Gap Acceptance Control System. In this system any intervehicular headway in the outside freeway lane in excess of T_1 seconds is defined as a gap into which a vehicle that seeks to enter the freeway may be placed. To achieve the placement of entering-vehicles into detected gaps, however, the gap detection process must be carried on sufficiently upstream from the merge point to allow both for synchronization after suitable travel time of vehicles on the ramp and for survival of detected gaps during the passage from the detection point to the merge point. In addition, the gap threshold T_1 must be sufficiently large so that an acceptable portion of the drivers execute a synchronized moving merge. Within this framework the design problem is then to select locations for the traffic control signal and the gap detector equipment and to specify the gap threshold T_1.

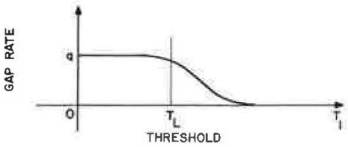

Figure 1. Gap detection rate as a function of threshold limit and lane volume.

Location of the control signal is restricted by the need for sufficient room to accelerate prior to the merge maneuver and by the desire to provide sufficient room for the queue that develops behind the signal. Coordinated with this is the need to place the gap detector as close to the merge point as possible to obtain the best possible gap survival conditions. The choice at any particular ramp is thus limited. The specification of T_1 to provide an acceptable rate of gap acceptance by merging drivers, however, is arbitrary and a large range for choice exists. When T_1 is chosen to be small, there are many gaps and the smaller of these are likely to be rejected by the drivers. For larger values of T_1 there are fewer gaps, but the smaller of the gaps are deleted and the likelihood that a driver will reject the offered gap is reduced.

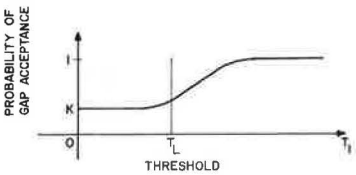

Figure 2. Gap acceptance by driver as a function of gap size.

If the design concept is to be properly established, first consideration must be given to the rate at which gaps (i.e., headways in excess of T_1) appear in a traffic stream of volume q. Figure 1 shows that the gap detection rate, $\mu(T_1, q)$, is a function of both the threshold limit T_1 and the lane volume q. Then, it is necessary to note that for very large gaps the probability that a driver accepts the gap for a moving merge is approximately 1, and for sequentially smaller gaps the probability of acceptance by the driver is reduced toward a finite limit, K, between 0 and 1 (Fig. 2). This lower limit corresponds to the probability that a driver can force entrance into a stream at a point at which no headway was detected. Based on these 2 concepts, subdivision of all gaps into 2 groups is then possible; the 2 groups include those that are accepted for a moving merge and those that are not accepted for a moving

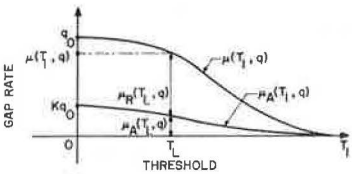

Figure 3. Gaps accepted and not accepted by drivers for moving merge.

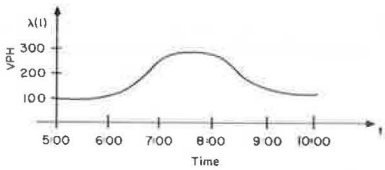

Figure 4. Mean of time-dependent Poisson process used to characterize the ramp arrival process.

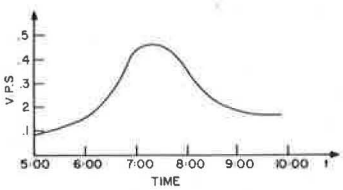

Figure 5. Time-dependent freeway volume.

merge (Fig. 3). Corresponding to this division, $\mu_A(T_1, q)$ is the rate at which accepted gaps are detected, and $\mu_R(T_1, q)$ is the rate at which rejected gaps are detected; the sum of these quantities, $\mu(T_1, q)$ is the rate at which all gaps in excess of T_1 are detected, that is,

$$\mu(T_1, q) = \mu_R(T_1, q) + \mu_A(T_1, q) \tag{1}$$

Single-Vehicle-Release-After-Completion Mode

For any specified value of gap threshold, $T_1 = T$, the rate at which gaps are detected is simply $\mu(T, q)$. Thus a driver who randomly arrives at the controller suffers an expected delay of $1/\mu(T, q)$ seconds before a gap is detected. Then, the driver must proceed down the ramp for which the expected travel time is denoted by t_r. In addition, the probability that the offered gap is unacceptable is equal to the ratio of unacceptable gaps to total gaps, $\mu_R(T, q)/\mu(T, q)$. Hence, an extra expected delay associated with the rejection of an offered gap is

$$\frac{\mu_R(T, q)}{\mu(T, q)} \cdot \frac{1}{\mu_S(q)}$$

where $\mu_S(q)$ is the occurrence rate of gaps that are acceptable for merge by drivers who have previously rejected the gap offered to them and are stopped in the merge zone. These quantities can be used to evaluate the total expected service time for a driver who arrives at the ramp control signal for merge service onto the freeway as

$$t_e = \frac{1}{\mu(T, q)} + t_r + \frac{\mu_R(T, q)}{\mu(T, q)} \cdot \frac{1}{\mu_S(q)} \tag{2}$$

As a result, when operation is restricted to a one-at-a-time mode and release is predicated on the completion of the merge maneuver by all previously released vehicles, a single-vehicle-release-after-completion system exists for which the ramp service rate μ' is a function of the threshold limit T and the volume q in the outside freeway lane. This service rate is

$$\mu' = \frac{1}{t_e} = \frac{\mu_S(q)\,\mu(T, q)}{\mu_R(T, q) + \mu_S(q)\left[1 + t_r\,\mu(T, q)\right]} \tag{3}$$

when t_e is large compared with $1/q$ so that the time interval between the instants at which sequential gaps are sought is large compared with average headway in the stream. (Note that the condition probability of a gap at the instant $t = T_0$ given a gap at the instant $t = 0$ approaches the unconditioned probability of a gap at the instant $t = T$ when T_0 is made very large. For an Erlang headway process on the freeway, this is equivalent to $t_e \gg 1/q$.)

The Design of a Controller for the Single-Vehicle-Release Mode

One useful and fairly common model for an urban freeway employs stochastic processes with slowly varying parameters to account for both ramp arrivals and highway flows. In particular, the peak-period ramp-arrival process is often described as a time-dependent Poisson process (1) with $\lambda(t)$ as shown in Figure 4. In addition, the

intervehicular spacings for vehicles on the highway are described as independent samples from an Erlang distribution (2, ch. 9):

$$f(t) = \frac{(aq)^a \, t^{a-1} \, e^{-aqt}}{(a-1)!} \tag{4}$$

where a is an integer and q varies slowly as a function of time corresponding to average volume (Fig. 5). Based on these descriptions and subject to the assumption that the process parameters vary slowly, one controller design philosophy is to maximize the service rate μ' subject to a limitation of downstream freeway capacity. The service rate μ' can be rewritten with the arguments omitted as

$$\mu' = \frac{1}{t_r + (\mu_R + \mu_S)/\mu\mu_S} = \frac{1}{t_r + 1/\mu''}$$

From this form it is seen that, for a given t_r, the maxima of μ' correspond to the maxima of μ''. Hence, setting $d\mu''/dT = 0$ located the values of T corresponding to the relative minima of μ'. In particular,

$$d\mu''/dT = \frac{\mu_S(\mu_R + \mu_S)\,(d\mu/dT) - \mu_S\mu(d\mu_R/dT)}{(\mu_R + \mu_S)^2} = 0 \tag{5}$$

reduces to

$$(\mu_R + \mu_S)\,(d\mu/dT) = (d\mu_R/dT)\,\mu \tag{6}$$

because $(\mu_R + \mu_S)^2$ is positive and finite, and $\mu_S \neq 0$. From Eq. 6,

$$(d/dT)\,[\ln(\mu) - \ln(\mu_R + \mu_S)] = 0 \tag{7}$$

Because T corresponds to the threshold that is set on the minimum spacing between vehicles on the freeway into which a moving merge will be attempted, all spacings in excess of this threshold are gaps. For any given threshold limit, the rate at which gaps appear in a stream with volume equal to q vehicles per second is

$$\mu = q \int_T^{\infty} \frac{(aq)^a \, t^{a-1} \, e^{-aqt} \, dt}{(a-1)!}$$

$$= q\, e^{-aqT} \sum_{i=0}^{a-1} \frac{(aqT)^i}{i!} \tag{8}$$

When the probability that a presented gap of t is accepted by a driver for a moving merge is described (2, ch. 9) by

$$P_a(t) = 1 - e^{-Kt} \tag{9}$$

then the rate of successful moving merges is

$$\mu_A = q \int_T^{\infty} (1 - e^{-Kt}) \left[\frac{(aq)^a \, t^{a-1} \, e^{-aqt}}{(a-1)!}\right] dt$$

$$= q \left\{ e^{-aqT} \sum_{i=0}^{a-1} \frac{(aqT)^i}{i!} - \frac{(aq)^a}{(aq+K)^a} \, e^{-(aq+K)T} \sum_{i=0}^{a-1} \frac{[(aq+K)T]^i}{i!} \right\} \tag{10}$$

and the rate of unsuccessful moving merges is

$$\mu_R = q \frac{(aq)^a}{(aq + K)^a} e^{-(aq+K)T} \sum_{n=0}^{a-1} \frac{[(aq + K)T]^n}{n!} \tag{11}$$

When the indicated derivatives are evaluated and employed in Eq. 6, then algebraic manipulations yield

$$\mu_S = \frac{q\, e^{-(aq+K)T}}{(aq + K)^a} \left\{ \sum_{n=0}^{a-1} \frac{[(aq + K)^a (aq)^n - (aq + K)^n (aq)^a]\, T^n}{n!} \right\} \tag{12}$$

as the equation from which the optimum threshold settings are obtained, subject to the constraint $T \geq 0$. Because this equation is of the form $\mu_S = f(T)$ where

$$f(T) = A\, e^{-BT} \sum_{n=0}^{a-1} C_n\, T^n \tag{13}$$

and because

$$C_n > 0 \text{ for every } n \tag{14}$$

the sum

$$\sum_{n=0}^{a-1} C_n\, T^n > 0 \text{ for } T > 0 \tag{15}$$

Thus, $f(T) > 0$ for $T > 0$. In addition, the derivative of the expression $f(T)$ is

$$f'(T) = (d/dT) \left[A\, e^{-BT} \sum_{n=0}^{a-1} C_n\, T^n \right] \tag{16}$$

where

$$f'(T) = -KA(aq + K)^a \left[e^{-BT} \sum_{n=0}^{a-1} \frac{(aqT)^n}{n!} \right] \tag{17}$$

Inspection of this quantity reveals that $f'(T)$ is negative for all positive T. Therefore, $f(T)$ is monotonic decreasing for $T \geq 0$. On this basis, the conclusion is that a unique optimum solution for the controller threshold T exists. That value is 0 when

$$\mu_S \geq q \frac{1}{(aq + K)^a} \left[(aq + K)^a - (aq)^a \right] \tag{18}$$

Otherwise, the value is the positive number T_{opt}, obtained as the solution to Eq. 12.

After T_{opt} is evaluated as described, the control policy is specified next. In particular, when the sum of the volume on the freeway and the demand on the ramp does not exceed the volume limitation for the freeway, a threshold of T_{opt} is used for the ramp. This ensures the highest possible service rate for demand on the ramp and results in minimum expected delay and minimum expected queue length. When the sum of the freeway volume and ramp demand exceeds the volume limitation, a threshold of T_{opt} is not acceptable. Instead a threshold T_c must be employed such that the sum of the freeway volume and the served portion of the demand is equal to or less than the volume limitation.

At this point, the heretofore undefined quantities of freeway volume, ramp service, and volume limitation must be considered more exactly. For this purpose it is noted that experience and traffic flow theory (3) indicate that the short-term production of any given point on a freeway cannot exceed some upper bound Q, approximately 2,000 vehicles per hour (vph), without significant risk of breakdown in the flow of traffic. Thus the T_A minute running average of flow past a critical point must be limited to approximately QT_A vehicles, depending on the facility. When this short-term average volume is below the specified capacity limit, additional vehicles from the ramp can be admitted into the stream, provided that the running average of the sum remains below the critical value. Thus the average ramp service is limited at most to the difference between the limiting and actual average volumes.

Many solutions are possible to the controller design problem that satisfy this restriction. One such possibility allows the threshold to be set at T_{opt}, provided that the restriction is met, and inhibits all merging otherwise. By this process all available roadway capacity is used as quickly as possible, and additional waiting vehicles are inserted into gaps as additional room for single vehicles arises. This solution, when associated with the limit condition in which T_{opt} equals 0, becomes the capacity adjusted metering system (4, 5).

A second approach to the controller design problem involves the gradual adjustment of the threshold as a function of freeway volume. This technique has the usual advantages associated with smooth variation in controller policy; however, it also has the added limitation, extra delay, associated with smoothing. In particular, the first policy carries the risk of inserting an acceptable averaged number of vehicles into the stream too quickly thus causing breakdown, and the second policy includes the risk of adjusting too slowly thus overloading the stream.

The Multiple-Vehicle-Per-Gap Merge Mode

In the generalized release-after-completion merge mode, a string of vehicles is released to attempt to merge into a suitable single gap after all previously released vehicles have completed the merge operation. To design a controller for this action requires only the specification of the gap threshold T_n that corresponds to the smallest gap into which a string of n vehicles may attempt a merge. When this is done, the sequence of numbers $\{T_n\}$ determines the number of vehicles that may attempt to merge into any given gap.

The subsequent analysis and controller design is simplified by the assumption that the gap threshold is chosen so that the probability that the last vehicle in a released string of n fails to merge is substantially larger than the probability that 2 or more vehicles in the string balk. Thus, the expected service time t_{en} for a string of n vehicles that is released to attempt to merge into a gap in excess of T_n but less than T_{n+1} is

t_{en} = the expected travel time for the first vehicle to reach the merge zone + the sum of the expected intervehicular headways between vehicles in the string + the expected extra delay due to gap rejection by the n th vehicle

$$= t_r + (n - 1)\,h + (\text{Prob n are released and the n th balks})\,(1/\mu_S)$$

$$= t_r + (n - 1)\,h + (1/\mu_S)\,\frac{[\mu_{Rn}(T_n) - \mu_{Rn}(T_{n+1})]}{[\mu(T_n) - \mu(T_{n+1})]} \tag{19}$$

where

t_r = the expected travel time for 1 vehicle on the ramp,
h = the expected intervehicular headway,
μ_S = the rate at which vehicles stopped at the merge zone execute merges into the freeway stream,
$\mu(T)$ = the rate at which gaps in excess of the threshold setting of T appear in the stream, and
$\mu_{Rn}(T)$ = the merge rejection rate for the n th vehicle in a string when the release threshold is set at T.

Next, it is noted that the rate at which merge opportunities appear in the stream, μ_{Total}, is

$$\mu_{Total} = \sum_{n=1}^{\infty} n \text{ (the rate of merge opportunities for strings of exactly n vehicles)}$$

$$= \sum_{n=1}^{\infty} n[\mu(T_n) - \mu(T_{n+1})] = \sum_{n=1}^{\infty} \mu(T_n) \tag{20}$$

Based on this, the expected service time associated with the merge opportunities offered by the stream T_e is

$$T_e = \sum_{n=1}^{\infty} \left[\begin{array}{l}\text{(merge opportunity rate for strings of n vehicles) (total expected delay}\\ \text{for the merge of a string of n vehicles)/(total merger opportunity rate)}\end{array}\right]$$

$$= \sum_{n=1}^{\infty} \frac{\mu(T_n) - \mu(T_{n+1})}{\mu_{Total}} \left\{ t_r + (n-1)h + (1/\mu_S)\left[\frac{\mu_{Rn}(T_n) - \mu_{Rn}(T_{n+1})}{\mu(T_n) - \mu(T_{n+1})}\right] + \frac{1}{\mu(T_1)} \right\}$$

$$= \frac{1}{\mu_{Total}} \left\{ (t_r - h)\mu(T_1) + h\mu_{Total} + (1/\mu_S) \sum_{n=1}^{\infty} [\mu_{Rn}(T_n) - \mu_{Rn}(T_{n+1})] + 1 \right\}$$

$$= h + \frac{\sum_{n=1}^{\infty} [\mu_{Rn}(T_n) - \mu_{Rn}(T_{n+1})] + [1 + (t_r - h)\,\mu(T_1)]\,\mu_S}{\mu_S \mu_{Total}} \tag{21}$$

Finally, $\mu_{Rn}(T_n) >> \mu_{Rn}(T_{n+1})$ is implied by the assumption that the probability of gap rejection by the last released vehicle in a string is much greater than the probability of a balk by any other vehicle in that string. Thus,

$$T_e = h + \frac{\sum_{n=1}^{\infty} \mu_{Rn}(T_n) + \mu_S\,[1 + (t_r - h)\,\mu(T_1)]}{\mu_S \mu_{Total}} \tag{22}$$

and the merge capacity limit for the stream is

$$\mu_{Stream} = \frac{1}{T_e} = \frac{1}{h + \dfrac{\mu_S \mu_{Total}}{\sum_{n=1}^{\infty} \mu_{Rn}(T_n) + \mu_S[1 + (t_r - h)\mu(T_1)]}}$$

$$= \frac{1}{(h + t')} \tag{23}$$

where

$$t' = \frac{\mu_S \mu_{Total}}{\sum_{n=1}^{\infty} \mu_{Rn}(T_n) + \mu_S[1 + (t_r - h)\mu(T_1)]} \tag{24}$$

Hence, the merge service capacity is maximized when t' is minimized. A controller design based on this is considered next.

The Design of a Controller for String-Release Mode

The extension of controller action to include the release of strings of vehicles for insertion into adequately large gaps offers several advantages at the expense of separate, individualized merge service. In particular, the safety level associated with the single-vehicle-release-after-merge mode is exchanged for a higher merge capacity when demand indicates that this is necessary. The result is a control law that offers individual service when demand is low and that allows for extended merge capability when demand is high, provided that adequate freeway capacity exists.

Inspection of the expression for the capacity of the ramp controller, μ_{Stream}, reveals this quantity depends on the set of threshold settings $\{T_n\}$. In general, this expression need not be unimodal, and the coordinates of the local maxima are not always identified by a sequential search. There is a good basis, however, for setting the thresholds by a sequential optimization. In particular, because there is no information included in the model to indicate the actual demand for service, an increase in the ramp capacity by the increase of the 2-vehicle merge rate at the expense of the 1-vehicle merge rate is detrimental to system performance when the actual demand requires only 1-vehicle merges. This situation is contrasted by the following: When the value of T_1 is selected to maximize the 1-vehicle merge process and then T_2 is selected to maximize the 2-vehicle merge process given T_1, the resultant system always provides the highest 1-vehicle merge rate and yields extra capacity through 2-vehicle merges as required by demand. Although this controller may not provide as large a total ramp capacity as is available by the simultaneous selection of T_1 and T_2, only the actual demand will determine under which controller the ramp provides better service.

When the intervehicular headways on the freeway are described as independent samples from the Erlang distribution

$$f(t) = \frac{(aq)^a \, t^{a-1}}{(a-1)!} \, e^{-aqt} \tag{25}$$

and the probability that the last vehicle in a string of n vehicles accepts an offered gap of t seconds or less is described by the cumulative Erlang

$$P_{an}(t) = 1 - e^{-S_n b_n t} \sum_{m=0}^{S_n - 1} \frac{(S_n b_n t)^m}{m!} \tag{26}$$

the values of $\mu_{Rn}(T_n)$, $\mu(T_n)$, and μ_S can be evaluated. Here

$$\mu(T_n) = q \int_{T_n}^{\infty} f(t)dt = qe^{-aqT_n} \sum_{\ell=0}^{a-1} \frac{(aqT_n)^\ell}{\ell!} \tag{27}$$

and

$$\mu_{Rn}(T_n) = q \int_{T_n}^{\infty} [1 - P_{an}(t)] \, f(t)dt = q \sum_{m=0}^{s_n - 1} \left[\frac{(s_n b_n)^m}{m!} \, \frac{(aq)^a}{(a-1)!} \, \frac{(a+m-1)!}{(aq + s_n b_n)^{a+m}} \right]$$

$$\cdot \left[e^{-(aq+s_n b_n)T_n} \sum_{\ell=0}^{m+\ell-1} \frac{(aq + s_n b_n)^\ell T^\ell{}_n}{\ell!} \right] \tag{28}$$

Likewise, when $P_s(t)$ is the standing merge headway acceptance probability,

$$\mu_S = q \int_0^\infty P_s(t)\, f(t)dt$$

and

$$\mu_S = q \sum_{m=0}^{r-1} \frac{(a + m - 1)!\ (rc/aq)^m}{m!\ (a - 1)!\ (1 + rc/aq)^{a + m}} \tag{29}$$

when $P_s(t)$ is cumulative Erlang type r with mean equal to 1/C.

The expression for ramp capacity for this case has been evaluated for some examples subject to the philosophy of sequential optimization. Capacity versus threshold is shown for outside lane volumes between 800 and 2,000 vehicles per hour in Figures 6 through 13 to indicate the system sensitivity to adjustment of the threshold T_1. Figures 14 through 21 show capacity versus string length for similar volumes. In all cases the lower order thresholds are set at their optimum value before the threshold of interest is varied. In addition, it is assumed that the gap acceptance probability that corresponds to the end vehicle in a string is type 3 Erlang with a mean that varies linearly with string length.

Finally, it is noted that when the maximum string length is restricted to 1, and the gap acceptance probability for single vehicles is Erlang type 1, the string merge capacity simplifies into the 1-vehicle merge capacity previously developed.

CAPACITY AND SENSITIVITY OF A STRING-RELEASE-AFTER-COMPLETION MERGE-CONTROLLER—AN EXAMPLE

At this point the capacity of a ramp control system that releases a string of vehicles to attempt a merge into the freeway stream is considered. For this operational mode the number of vehicles in the string is dependent on the size of the detected gap, and release is conditional on the completion of successful merges by all previously released vehicles. Hence, the ramp is clear before an additional string is released.

Several cases considered for the controller design presented in the previous section included the following:

1. Two alternate length entrance ramps were measured in "seconds of time between the instant the control signal is turned to green and the instant the first merge-vehicle arrives at the merge zone"; thus $t_r = 5$ seconds and $t_r = 10$ seconds are used to investigate the effect of ramp length.
2. Four alternate gap acceptance criteria were employed to describe drivers that balk at the moving-merge attempt. Here 1/c equal to 1.5, 3.0, 4.5, and 6.0 seconds are respectively used for the mean gap sizes that drivers who stop at the merge zone find acceptable for a standing merge.
3. Five freeway volumes that correspond to q equal to 800, 1,100, 1,400, 1,700, and 2,000 vph were investigated to study the effects of variation in volume.

In addition, the following parameter values were employed as reasonable or typical or both:

1. The standing-merge gap acceptance probability is described as Erlang type 3, thus r = 3;
2. The moving-merge gap acceptance probability for the n th vehicle in a string is Erlang type 3, thus $s_n = 3$ for all n;
3. The mean acceptable gap for the n th vehicle in a string is $1/b_n = 3n - 1.5$, thus the first, second, third, and n th vehicles accept a gap of 1.5, 4.5, 7.5, and 3n - 1.5 seconds respectively on one-half of the merge attempts; and
4. The freeway intervehicular headway process is described by an Erlang type a distribution with mean equal to 1/q (where q is the volume), and here, a is the nearest integer value of the expression $a = 0.92\ e^{3.6q}$ and is based on the experimental results obtained on the Gulf Freeway (2).

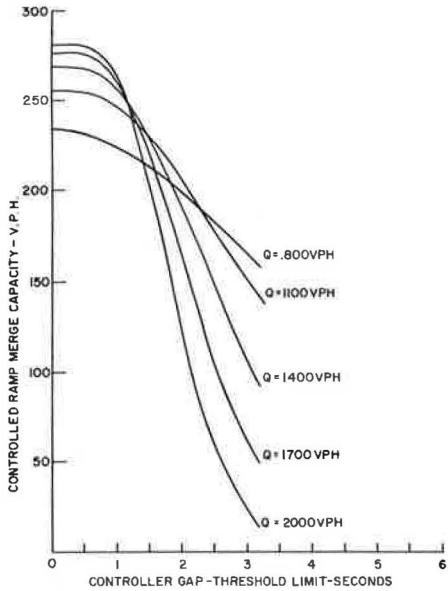

Figure 6. Single-vehicle-release-after-completion mode with expected ramp travel time of 10 seconds, mean acceptable gap for moving merge of 1.5 seconds, and mean acceptable gap for standing merge of 1.5 seconds.

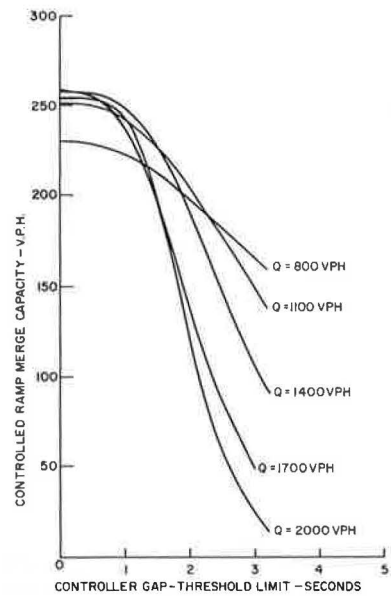

Figure 7. Single-vehicle-release-after-completion mode with expected ramp travel time of 10 seconds, mean acceptable gap for moving merge of 1.5 seconds, and mean acceptable gap for standing merge of 3.0 seconds.

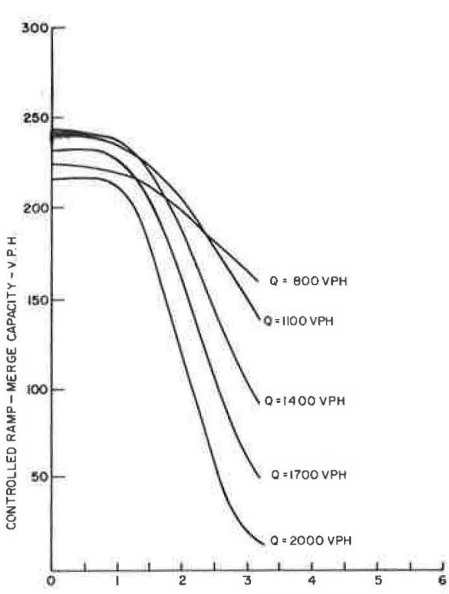

Figure 8. Single-vehicle-release-after-completion mode with expected ramp travel time of 10 seconds, mean acceptable gap for moving merge of 1.5 seconds, and mean acceptable gap for standing merge of 4.5 seconds.

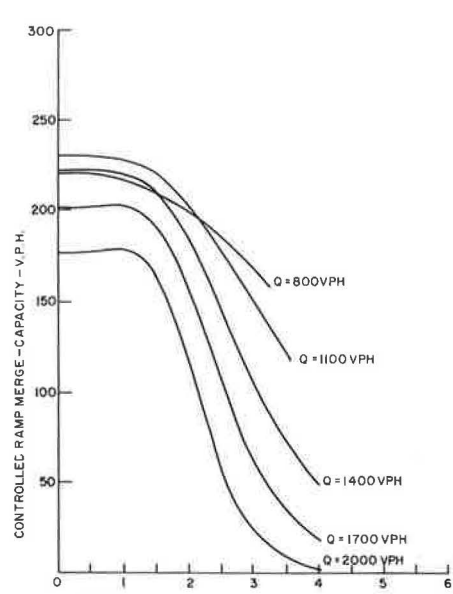

Figure 9. Single-vehicle-release-after-completion mode with expected ramp travel time of 10 seconds, mean acceptable gap for moving merge of 1.5 seconds, and mean acceptable gap for standing merge of 6.0 seconds.

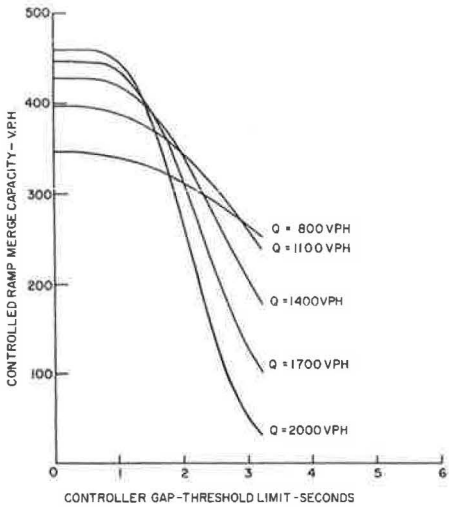

Figure 10. Single-vehicle-release-after-completion mode with expected ramp travel time of 5 seconds, mean acceptable gap for moving merge of 1.5 seconds, and mean acceptable gap for standing merge of 1.5 seconds.

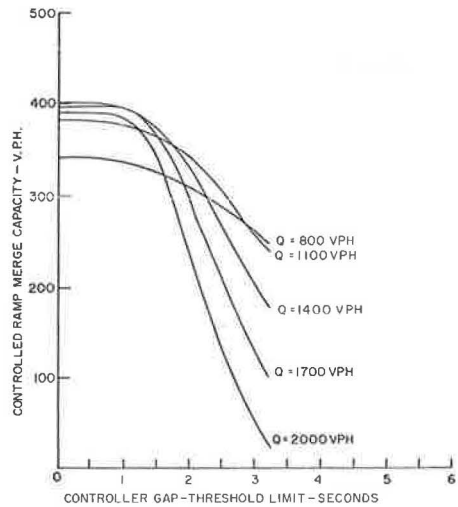

Figure 11. Single-vehicle-release-after-completion mode with expected ramp travel time of 5 seconds, mean acceptable gap for moving merge of 1.5 seconds, and mean acceptable gap for standing merge of 3.0 seconds.

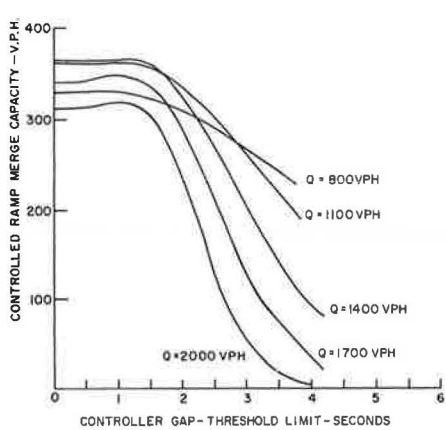

Figure 12. Single-vehicle-release-after-completion mode with expected ramp travel time of 5 seconds, mean acceptable gap for moving merge of 1.5 seconds, and mean acceptable gap for standing merge of 4.5 seconds.

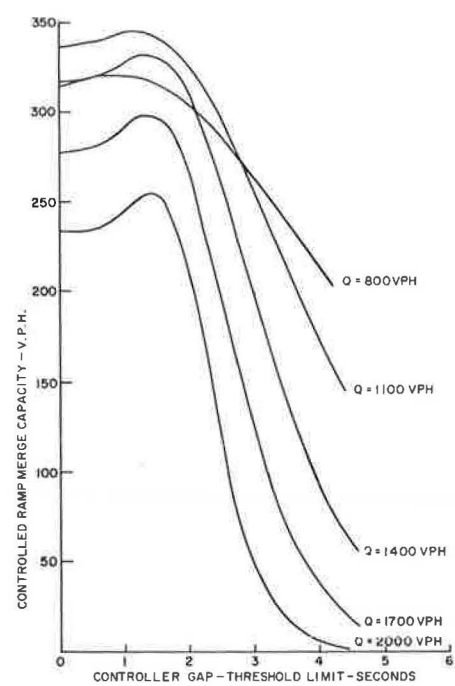

Figure 13. Single-vehicle-release-after-completion mode with expected ramp travel time of 5 seconds, mean acceptable gap for moving merge of 1.5 seconds, and mean acceptable gap for standing merge of 6.0 seconds.

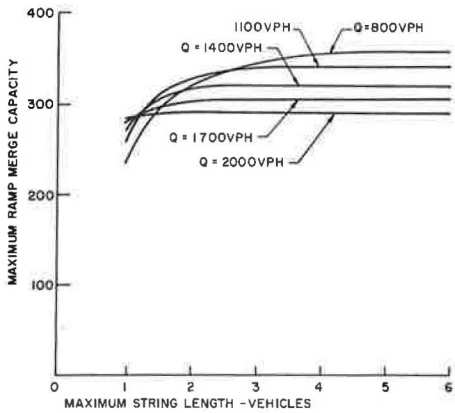

Figure 14. Multiple-vehicle-release-after-completion mode with expected ramp travel time of 10 seconds, mean acceptable gap for moving merge for n th vehicle of 3n - 1.5 seconds, and mean acceptable gap for standing merge of 1.5 seconds.

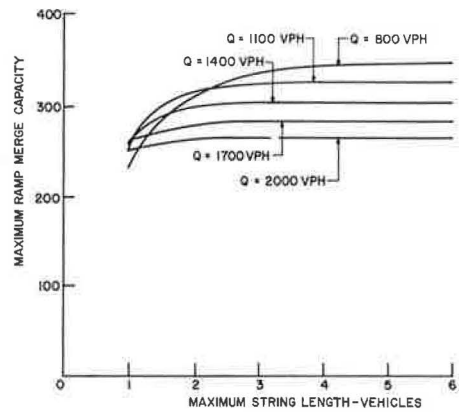

Figure 15. Multiple-vehicle-release-after-completion mode with expected ramp travel time of 10 seconds, mean acceptable gap for moving merge for n th vehicle of 3n - 1.5 seconds, and mean acceptable gap for standing merge of 3.0 seconds.

Before the results of the numerical analysis are discussed, 2 points are worth noting

1. When the time required by a vehicle to reach the merge zone is t_r seconds (after the signal is changed to green), then the maximum ramp volume in the single-vehicle-release-after-completion mode is $1/t_r$. This quantity equals 720 and 360 vph when the respective values of t_r are 5 and 10 seconds. When strings of n vehicles each are released in the release-after-completion mode and the assumption is used that intervehicular headway is n equal to 3 seconds, the corresponding merge rate is $1/(t_r + 3n - 3)$.

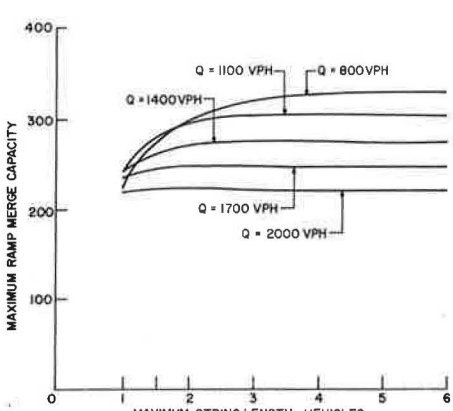

Figure 16. Multiple-vehicle-release-after-completion mode with expected ramp travel time of 10 seconds, mean acceptable gap for moving merge for n th vehicle of 3n - 1.5 seconds, and mean acceptable gap for standing merge of 4.5 seconds.

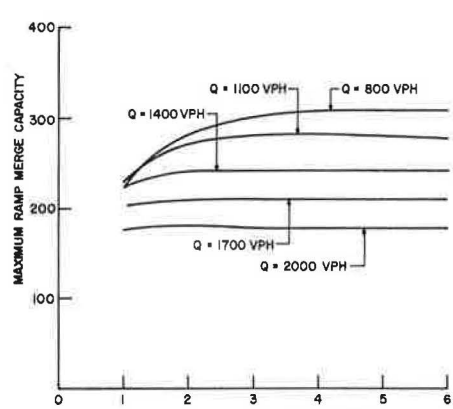

Figure 17. Multiple-vehicle-release-after-completion mode with expected ramp travel time of 10 seconds, mean acceptable gap for moving merge for n th vehicle of 3n - 1.5 seconds, and mean acceptable gap for standing merge of 6.0 seconds.

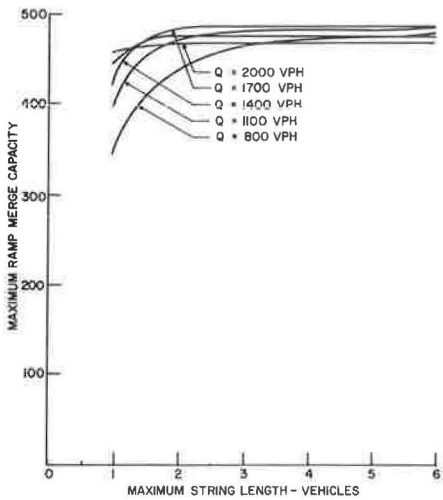

Figure 18. Multiple-vehicle-release-after-completion mode with expected ramp travel time of 5 seconds, mean acceptable gap for moving merge for n th vehicle of 3n - 1.5 seconds, and mean acceptable gap for standing merge of 1.5 seconds.

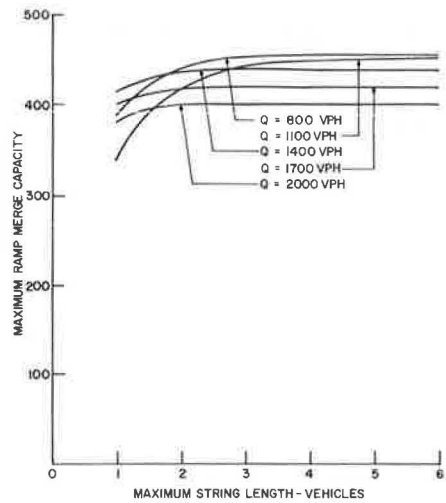

Figure 19. Multiple-vehicle-release-after-completion mode with expected ramp travel time of 5 seconds, mean acceptable gap for moving merge for n th vehicle of 3n - 1.5 seconds, and mean acceptable gap for standing merge of 3.0 seconds.

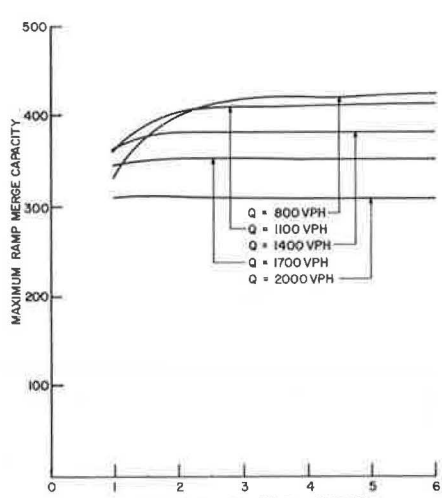

Figure 20. Multiple-vehicle-release-after-completion mode with expected ramp travel time of 5 seconds, mean acceptable gap for moving merge for n th vehicle of 3n - 1.5 seconds, and mean acceptable gap for standing merge of 4.5 seconds.

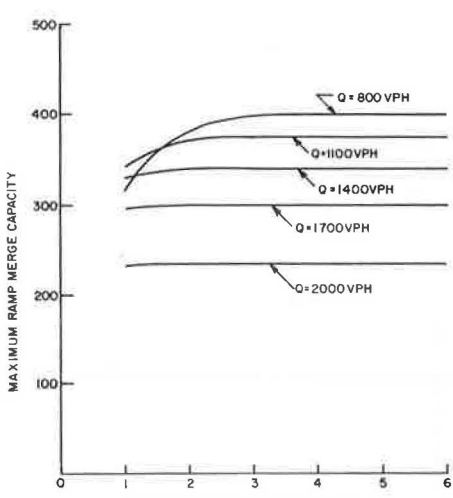

Figure 21. Multiple-vehicle-release-after-completion mode with expected ramp travel time of 5 seconds, mean acceptable gap for moving merge for n th vehicle of 3n - 1.5 seconds, and mean acceptable gap for standing merge of 6.0 seconds.

As n approaches infinity, this quantity approaches 1,200 vph for all values of t_r. This limit corresponds to the uncontrolled capacity limit of the entrance ramp.

2. The capacity of an entrance ramp in the release-after-completion mode depends on both the expected length and the expected frequency of the strings of vehicles that may be released to attempt a moving-merge into the freeway stream. For such operation, an increase in the freeway volume always results in a decrease in the number of vehicles in a released string. The expected number of released strings (i.e., the expected frequency), however, may either increase or decrease with an increase in freeway volume because this quantity depends on both the number and the suitability of freeway gaps. In particular, for low freeway volumes almost all gaps are suitable for a merge attempt by 1 vehicle, and an increase in the number of gaps implies an increase in the allowable number of released strings; however, for high freeway volumes only a portion of the gaps are suitable for merge attempts, and an increase in the number of gaps implies both that the average gap becomes smaller and that the number of gaps suitable for merge attempts decreases.

The use of parameter values previously listed in the models that were developed in the prior section makes it possible to numerically evaluate representative entrance ramp capacities. The results of such numerical evaluation with a string-release-after-completion merge controller are presented graphically for ease of interpretation. In Figures 6 through 13, the ramp capacity is shown as a function of threshold setting for the single-vehicle-release-after-completion mode, with volume as a parameter. These curves show that a threshold of approximately 1 second yields a generally near-maximum capacity for all cases. However, the sensitivity of capacity to threshold setting varies substantially. The short ramps (i.e., t_r = 5 seconds) tend to have higher capacities and greater sensitivity; the t_r = 10 second cases have lower rates and are less sensitive to variations in threshold. Likewise, there is a functional dependence of ramp capacity on the nature of the merge zone. For short ramps, the ease of execution of standing merges substantially affects the ramp capacity, but on long ramps this quantity is less significant.

More exactly, the numerically evaluated results indicate the following:

1. As the threshold is increased the occurrence of balks becomes less significant, and the capacities are asymptotically identical (i.e., independent of c for a given q and t_r).

2. The quality of service is improved at the expense of added delay when the threshold is increased, and the expected service time for a given threshold T_e is simply the reciprocal of the associated ramp capacity.

3. For all cases investigated, the largest ramp capacity for a single-vehicle-release-after-completion controller corresponds to a short ramp with an excellent merge zone (i.e., t_r = 5 and 1/c = 1.5 seconds). Here, a ramp capacity of approximately 450 vph is indicated for a heavy freeway flow condition (i.e., q of 1,700 to 2,000 vph), but actual service must be suitably restricted below this value to prevent a breakdown of the stream.

4. The differences between the ramp capacities shown by the curves in Figures 6 through 13 and the maximum capacities of 360 and 720 vph that correspond to t_r = 10 and 5 seconds respectively are attributed to the combined effects of limiting by the threshold setting and by gap rejection (i.e., by both moving and stopped vehicles).

The maximum ramp capacity available in the string-release-after-completion mode has been evaluated by a sequential search, subject to the restriction that a second vehicle is not released unless a gap of 3 seconds or larger is detected (i.e., because the headway h = 3 seconds). Similarly, the release of a third vehicle required a gap of at least 6 seconds, the release of a fourth vehicle required a minimum gap of 9 seconds, and so on. The resultant ramp capacities are shown as a function of maximum string lengths in Figures 14 through 21 with freeway volume as a parameter and for various values of t_r and c. These curves show that the largest benefit of multiple vehicle release occurs at low freeway volumes; at high volumes, little extra capacity is gained. Hence, the largest volume available for the cases investigated is approximately 480 vph when t_r = 5 seconds, 1/c = 1.5 seconds, and q = 800 vph. This is substantially below the limit of 1,200 vph that corresponds to q = 0. However, when the case of t_r = 10 seconds and 1/c = 3 seconds is considered as an example, it is observed that, when the freeway volume gradually drops from 2,000 to 800 vph, the single-vehicle-merge ramp

capacity remains almost constant at approximately 250 vph and the string-release mode yields an automatic increase in capacity up to 350 vph. Finally, it is noted that, when a ramp capacity of 400 vph is required for this ramp, a reduction in t_r from 10 seconds to 5 seconds yields a string-release capacity in excess of 400 vph. However, when a ramp capacity of 500 vph is needed, use of the string-release-after-completion mode is unacceptable and release-before-completion must be implemented. This mode of operation is considered in the next section.

One final point remaining with regard to the numerical results is that the parameter values were chosen on the basis of reasonableness in order to yield typical results for presentation. The models developed are quite general and may be applied to alternate situations by the proper selection of parameter values.

THE RELEASE-BEFORE-COMPLETION MODE

Although the release-after-completion mode exhibits many desirable attributes, such as high safety level and high quality of merge service, the question of whether to release an additional vehicle while previously released vehicles remain on the ramp requires consideration. In particular, when a possible merge opportunity arises for a vehicle awaiting service behind the ramp control signal, the expected service delay for that vehicle may be significantly reduced when that vehicle is released without awaiting merge-completion for the previously released vehicle. However, when this operational mode is to be employed, the question of what value of gap threshold to use must be reexamined.

When a possible merge opportunity is detected in the freeway traffic stream before all previously released vehicles have merged, the question of whether to release the next queued vehicle can be converted into a question of which action minimizes the expected delay. Here the expected delay subject to release of the vehicle must be compared with the expected delay subject to nonrelease to determine which decision to implement. The resultant analysis yields a controller that is a sequential decision-maker in which both the target gap sizes and present state of vehicles on the ramp affect the decision process.

Analysis of the operation of an entrance ramp in the release-before-completion mode can take several forms. One possibility is to assume that the actual ramp service rate is restricted to a value that does not cause a breakdown in the flow of vehicles on the freeway. Then, an investigation of expected delay for the next vehicle awaiting service simplifies to a question of which decision yields the earliest expected merge instant. For this case, the following points are noted:

1. There are n unmerged vehicles on the ramp;
2. The last previous release of a vehicle occurred at $t = t_2$;
3. The target gap sizes are $\tau_1, \tau_2, \ldots, \tau_n$ respectively;
4. A target gap of τ_{n+1} is detected at $T = t_d$;
5. The expected delay between the completion of standing merges is $1/\mu_S$;
6. The probability that the next vehicle can complete a moving merge into the prospective target is $\prod_{i=1}^{n+1} P_a(\tau_i)$; and
7. The expected delay before a successful merge, given the merge zone is cleared, is $t_r + 1/\mu''$ (i.e., μ'' is the merge rate when $t_r = 0$).

Based on these definitions, the expected additional time required to complete the merge of the (n + 1)th vehicle (i.e., beyond the time required to complete the merge of the nth vehicle and provided that no balk has occurred) is

$$\Delta t_y = \left[1 - \prod_{i=1}^{n+1} P_a(\tau_i)\right](1/\mu_S) + \left[\prod_{i=1}^{n} P_a(\tau_i)\right][t_d - t_\ell] \tag{30}$$

when the vehicle is released to attempt entry into the τ_{n+1} gap, and

$$\Delta t_n = t_r + 1/\mu'' \tag{31}$$

when release is inhibited until the ramp is clear.

Comparison of these 2 quantities reveals that when the gap τ_{n+1} satisfies the inequality

$$P_a(\tau_{n+1}) \geq [(t_d - t_\ell)\mu_S - 1] \prod_{i=1}^{n} P_a(\tau_i) - [\mu_S t_r + (\mu_S/\mu'') - 1]$$

$$\geq f[\mu_S, \mu'', t_r, t_d, \tau_1, \tau_2, \ldots, \tau_n] \tag{32}$$

the (n + 1)th vehicle suffers less delay in the merge process by not waiting for the ramp to clear. Thus for the operational condition in which the function $f(\cdot)$ on the right side of the inequality is negative, the inequality is always satisfied, and any value of τ_{n+1} is acceptable for a merge attempt. Similarly, when $f(\cdot)$ exceeds unity, no gap is acceptable because $P_a(\tau_{n+1})$ is a gap-acceptance probability and cannot exceed unity. Finally, when $f(\cdot)$ is between 0 and 1, a threshold T_{n+1} exists such that

$$T_{n+1} = P_a^{-1}[f(\cdot)] \tag{33}$$

where $P_a^{-1}(\cdot)$ is the inverse of the gap acceptance probability function and $P_a^{-1}[P_a(t)] = t$.

For a ramp that has experienced a stoppage and offers zero probability of a moving merge for the (n + 1)th vehicle, the expected delays associated with the release or the nonrelease of that vehicle at t_d are respectively

$$\Delta t_y' = 1/\mu_S \tag{34}$$

and

$$\Delta t_n' = t_r + 1/\mu \tag{35}$$

Comparison of these quantities indicates that, for any ramp on which the expected travel time t_r exceeds the difference between the expected delay $1/\mu_S$ (associated with a standing merge from the merge zone) and the travel-free service time $1/\mu''$ (associated with the successful clear-ramp, single-vehicle-merge rate), the release of the next vehicle into any size gap is warranted. Setting the probability of a moving merge for the (n + 1)th vehicle in the previous case equal to 0 reinforces this conclusion; this is exactly the same criterion as that for release emerges.

The Design of a Controller for the Release-Before-Completion Mode

Under those operational conditions in which an increase becomes necessary or desirable in the production of a particular entrance ramp above the level achieved in the release-after-completion mode, it can be achieved by operation in the release-before-completion mode. To do so, however, involves a trade-off between the quality of merge-service offered and the merge production.

For a long ramp (i.e., t_r is large) with a merge zone that is unobstructed so that the expected standing merge delay is comparable to (though more than) the expected delay in the attempt of a moving merge, analysis has shown that the release-before-completion mode offers an improved merge rate for the ramp, provided that the freeway stream does not break down. This property allows for a variation of the controller threshold (or merge-rate limit or both) to increase the ramp production as required within the limits necessary to preserve the stability of the stream. Such operation has been implemented on the Gulf Freeway when queue lengths beyond the available storage capacity have made action necessary. In addition, ramp operation with a fixed threshold in the release-before-completion mode has been more the rule than the exception on that Freeway.

By contrast, when a ramp is short (i.e., t_r is small) and the merge zone is marginal so that standing merge opportunities are rare compared with moving merge opportunities,

operation in the release-before-completion mode does not always yield an increase in ramp merge capacity. In particular, analysis reveals that when

$$t_r < (1/\mu_S) - (1/\mu'') \tag{36}$$

a stoppage of a released vehicle warrants that the release-before-completion operation be inhibited until the ramp is cleared. The existence of this restriction in the release-before-completion operation, however, is temporary and in no way negates the overall advantage of increased ramp capacity that results, provided that the controller selects only adequately large gaps that occur within an adequately short period after the last previous release.

SUMMARY

Under either control philosophy described, the threshold that yields the maximum ramp capacity must be identified first and, based on that value, a threshold setting is established that both provides an appropriate level of merge service and preserves the stability of the freeway stream against breakdown in flow caused by excess density. By this process, merge capacity will often be reduced below its maximum value to increase the level of service offered to drivers. Thus, the controller may attempt to operate in a mode in which a single vehicle is released whenever the ramp is clear, the demand is low, and the probability of a successful moving merge is adequately high. When it is necessary to increase the available ramp merge capacity, the operational mode is modified so as to reduce the level of service by the relaxation of control limits until adequate merge capacity is achieved. In this way the controller is continuously adjustable to both actual demand and actual freeway volume.

Based on analysis of the results obtained by use of the control modes developed in this paper, we believe a sound analytic theory exists for the design of freeway entrance ramp control systems. In particular, the relationship between merge capacity and system parameters is explicitly developed for several control modes. These results make possible the evaluation of both the capacity limitations inherent in the design of a particular ramp control system and the control system parameters necessary to provide a specified merge capacity for the ramp.

To aid in the design process, several typical ramp control configurations were analyzed in detail. These included the following conditions: (a) placement of a traffic control signal at between 5 and 10 seconds of travel time from the merge zone, (b) mean gap acceptance limits for vehicles stopped in merge zone of between 1.5 and 6.0 seconds, and (c) freeway volumes in the outside lane of between 800 and 2,000 vph. As a result of this analysis, the following observations are made:

1. The gap rejection phenomenon was found to increase the minimum expected ramp service time, in the single-vehicle-release-after-completion mode, to a level between 1.25 and 3.0 times the expected vehicular travel time on that ramp. For example, for a long ramp with an expected travel time of 10 seconds, the maximum ramp capacity is reduced to between 50 and 80 percent of the 360-vph limit that the 10-second travel time imposes on this mode; 3,600 seconds per hour divided by a minimum of 10 seconds per vehicle yields a maximum flow in the single-vehicle-release-after-completion mode of 360 vph. Here, the larger capacity corresponds to a volume of 800 vph in the outside freeway lane; the lower capacity occurred with a freeway volume of 2,000. In both cases, the assumption is that a relatively good merge zone exists because the mean acceptable gap for a standing merge from the merge zone was defined to equal 1.5 seconds. For a second example, with an expected travel time of 5 seconds and a short merge zone, the results indicate a reduction in maximum ramp capacity to between 35 and 65 percent of the limit of 720 vph that the 5-second travel time imposes on this mode. Again, capacity depended on freeway volume.

2. The release of more than 1 vehicle per string can increase the ramp capacity without forsaking the release-after-completion operational mode. For such operation, multiple-vehicle-release opportunities are rare when high freeway volumes exist and

little is gained in these instances. However, when freeway volumes are low (e.g., 800 vph per lane), improvements of approximately 50 percent over single-vehicle operation are possible.

3. Operation in the string-release-after-completion mode appears to be limited at most to between 15 and 40 percent of the uncontrolled ramp capacity, with the higher production generally corresponding to the lower freeway volumes (i.e., 800 vph per lane). To provide ramp capacities in excess of this limit requires that release-before-completion operation be employed.

4. When the difference between the expected service time required to complete a standing merge from the merge zone is less than the total expected service time required to complete a moving merge from behind the ramp control signal (i.e., including the expected travel time), then the release-before-completion mode exhibits a maximum ramp capacity that equals the uncontrolled ramp capacity. However, the actual service provided by this controller must be constrained by consideration of freeway stream stability and actual demand characteristics. For example, metering at any fixed rate below 1,200 vph would be possible for several of the cases investigated in this report, provided that stream stability was not a problem.

REFERENCES

1. Highway Capacity Manual. HRB Spec. Rept. 87, 1965.
2. Drew, D. R. Traffic Flow Theory and Control. McGraw-Hill, New York, 1968.
3. Haight, F. A. Mathematical Models for Traffic Flow. Academic Press, New York, 1963.
4. Wattleworth, J. A. Peak-Period Control of Freeway Systems: Some Theoretical Considerations. Northwestern Univ., Evanston, Ill., PhD dissertation, 1963.
5. Hodgkins, E. A. Effects of Volume Controls on Freeway Traffic Flow—A Theoretical Analysis. Public Roads, Feb. 1968.
6. Buhr, J. H. Freeway Ramp Control. Civil Engineering, ASCE, June 1968.