Traffic Signal Delay Model for Nonuniform Arrivals

P. S. Olszewski

Platooning of arrivals has a significant effect on traffic signal performance and must be considered in estimating delay. The Highway Capacity Manual (HCM) method, which is based on arrival types, is not accurate because the progression adjustment factors vary in large steps. Another estimation method for pretimed signals is based on the step arrival flow approximation, which assumes different average arrival rates during the red and green periods of the cycle. The model requires only one easily measurable parameter—the proportion of arrivals during the red period. The delay estimates obtained using this method were in good agreement with cycle-by-cycle delay observations, except for cases in which the arrival rate profile was not symmetrical inside the red period. Compared with the HCM progression adjustment factors, the step arrival rate model generally predicts a stronger effect of platooning on uniform delay. The effect of platooning on overflow delay is not significant, however. Therefore, no adjustment factors should be used for the overflow delay component.

The delay vehicles experience at signalized intersections is one of the most important indicators for measuring intersection performance. Estimation of delay has become more common since the 1985 Highway Capacity Manual (HCM) (1) linked delay directly to the level of service such intersections provide.

Modeling of delay is complicated because of the large number of random and nonrandom factors affecting traffic arrival and departure. In a review of delay models developed so far, Hurdle (2) highlights their limitations and simplifying assumptions. For example, a common assumption is that the vehicle arrival rate is uniform throughout the signal cycle. This approximation is only reasonable if the arrival process is random (the expected value of vehicle headway is constant in time), that is, if the intersection is isolated from the effect of other traffic signals. In urban areas, however, almost every intersection is part of a large signalized network. Vehicles travel in dispersed platoons, and their arrival times at any given approach are largely determined by the timing of the upstream signal. If the two traffic signals are well coordinated, a vehicle platoon will arrive during the green period, with little delay incurred. With lack of proper coordination, most vehicles may arrive during the red period, and the delay will be considerably longer.

So far, the only method of calculating delay that takes platooned arrivals into account is given in the HCM (1). Adjustment factors were introduced to reflect the cases described above. This paper discusses an alternative method of calculating delay for pretimed (fixed-time) signals when the arrival rate is nonuniform.

BACKGROUND

Traffic signal delay has the following two components:

1. Uniform delay, incurred because some vehicles arrive during the red period of the cycle and have to wait their turns to depart when the signal turns green.
2. Overflow delay, which occurs because in some cycles, demand exceeds capacity for departures. This results from random fluctuations in the arrival flow rate or during prolonged periods of oversaturation.

The HCM (1) also distinguishes between these two delay components, although their meaning is explained somewhat differently (the overflow component is called “incremental delay”). The HCM (1) overflow delay formula has generated a lot of discussion (3–5), which will not be analyzed here because platooning of arrivals seems to have a much stronger effect on uniform delay.

The formula for calculating the uniform component of delay was developed as early as 1941. Its derivation has been explained by Hurdle (2), among others. The uniform delay \( d_u \) can be expressed as

\[
d_u = \frac{r^2}{2(t_c - gx)}, \tag{1}
\]

where

- \( r \) = effective red time,
- \( g \) = effective green time,
- \( t_c \) = cycle time, and
- \( x \) = degree of saturation or ratio of flow \((q)\) to capacity \((c)\); \( x = q/c \).

This formula is valid when the arrival rate is uniform throughout the cycle. It is used in the British (6) and Australian (7) delay calculation methods. Equation 1 is also the basis for delay calculation in the HCM (1), with two important modifications. The first uses stopped delay rather than total delay as a measure of performance for signalized intersections. Stopped delay is the time a vehicle is stationary. Total delay also includes the time lost because of deceleration and slower movement through the intersection. According to the HCM (1), total delay is approximately 1.3 times the stopped delay, so Equation 1 is used with the coefficient 1/1.3, or 0.76.

---

School of Civil and Structural Engineering, Nanyang Technological Institute, Nanyang Avenue, Singapore 2263.
The other correction in the HCM (1) reflects nonuniformity of arrival flow. Five arrival types are defined:

1. Dense platoon arrives at the beginning of the red period (most unfavorable case),
2. Dispersed platoon arrives during the red period,
3. Random arrivals,
4. Dispersed platoon arrives during the green period, and
5. Dense platoon arrives at the beginning of the green period (most favorable case).

To make identification of arrival types easier, an additional variable is introduced called the "platoon ratio" ($R_p$).

$$R_p = \frac{P_g}{u},$$

where

$$P_g = \text{proportion of vehicles arriving during the green period},$$

and $u = \text{the green time ratio, or the ratio of effective green time (g) to cycle time (t_c)}$.

The HCM (1) gives a range of platoon ratio values corresponding to each arrival type. In the case of random arrivals, a platoon ratio value close to 1.0 can be expected, because the proportion of vehicles arriving during the green period should be equal to the green time ratio.

The arrival type is used to find the value of the progression adjustment factor ($p_\gamma$). This factor is to be multiplied by the delay estimate obtained from Equation 1. Values of progression factors are tabulated and depend on arrival type, as well as degree of saturation. Part of the HCM (1) table concerning pretimed (fixed time) signals is presented as Table 1.

The estimation procedure outlined is not precise, because the progression adjustment factors vary in steps up to 40 percent. For example, consider two approaches with identical signal timing, a degree of saturation of 0.6, and very close platoon ratios of 0.5 and 0.51. According to Table 1, the first approach would be classified as Arrival Type 1 and the second as Type 2. If the delay calculated from Equation 1 was 40 sec, then the final adjusted delay estimates for the two approaches would be 54 and 74 sec—a difference of 20 sec.

So far, research into delay with nonuniform arrivals has been linked to the problem of designing effective signal progression. Dick (8) and later Staniewicz and Levinson (9) presented methods for delay calculation based on known characteristics of progression bandwidth. Roupahl (10) proposed a model that assumes two uniform arrival rates during each cycle: one within the progression bandwidth, and another outside the bandwidth. The model leads to a set of equations for determining delay when platoons arrive during green, red, or partly during green and partly during red periods. The equations Staniewicz and Levinson (9) and then Roupahl (10) proposed seem to predict delay more accurately than the HCM (1) model, but at the cost of introducing new variables such as the length of platoon (bandwidth) and the relative time when the leading vehicle arrives (offset). These variables are very difficult to measure in practice, especially when platoons are dispersed or when the arrival pattern is disrupted by turning movements.

### STEP ARRIVAL RATE MODEL

The step arrival rate model was adopted as a practical compromise between the need for greater precision than the HCM (1) model offered and the requirement that only parameters that are easy to measure should be used. One such parameter is the proportion of arrivals during the red period ($P_r$), which the HCM (1) method uses indirectly. It can be measured by simply counting the arrival volume during the red and green periods separately.

The proposed model is based on the assumption that vehicles arrive at a uniform rate ($q_g$) during the green period and at a different rate ($q_r$) during the red period. This step arrival rate profile is shown in Figure 1. In case of no overflow, the total delay all vehicles incur is equivalent to the area of the triangles $D_r$ and $D_g$. This area can be expressed in terms of the average flow ($q$) and the proportion of arrivals on red ($P_r$). The average delay per vehicle can now be found by dividing the area $D_r + D_g$ by the total number of arrivals per cycle, or $q_t$. This derivation, which was presented by Olszewski (11), leads to the following equation for average uniform delay with platooned arrivals:

$$d_{wp} = \frac{rP_g}{2} + \frac{gP_r^2}{2(1/x + P_r - 1)}$$

where

$$d_{wp} = \text{uniform delay with platooned arrivals, and}$$

$$P_r = \text{proportion of arrivals during the effective red period.}$$

In theory, Equation 3 can be used subject to the condition $P_r = (x - 1)/x$. However, this condition can only be violated when $x = 1$ and $P_r = 0$, in which case $d_{wp} = t_P/2 = 0$. When $x$ is greater than 1.0 (oversaturation), any increase in delay belongs to its overflow component, and the uniform component is constant. This means that when $x > 1$, then $x = 1$ should be substituted into Equations 1 or 3. The HCM (1) does not mention this important limitation. When the arrival rate is uniform throughout the cycle, that is, when $P_r = r/t_c$, Equation 3 reduces to Equation 1.

The step arrival rate model was first proposed by Messer et al. (12). The delay formula they presented is similar to Equation 3 and has been used in the PASSER family of computer programs for calculating arterial progression.
DELAY MODEL WITH OVERFLOW QUEUE

The delay model described so far is applicable only to cycles with no overflow queue. The delay given by Equation 3 reflects only the delay motorists experience during the same cycle in which they arrive (see Figure 1). However, if an overflow queue exists at the beginning of Cycle \( k \), part of the delay in that cycle is incurred by vehicles that arrived in Cycle \( k - 1 \). This is shown as area \( D2 \) in Figure 2. If not all vehicles from Cycle \( k \) are able to depart, some will incur extra delay waiting for the next green in Cycle \( k + 1 \). This extra delay corresponds to area \( D3 \) in Figure 2.

In case of overflow, a precise definition of average delay per vehicle becomes important. Some researchers use the definition, average delay incurred during Cycle \( k \) (or period \( T \)). This corresponds to area \( D1 \) in Figure 2, divided by the number of arrivals in Cycle \( k \). This average reflects the true delay experienced by motorists arriving during Cycle \( k \), only if the queue is exactly the same at the beginning and the end of Cycle \( k \). If a long queue exists at the beginning and only one motorist arrives, his delay will be much less than the total area \( D1 \). Thus, to reflect properly the level of service perceived by motorists arriving during Cycle \( k \), the average delay per vehicle should be calculated as

\[
d = \frac{D1 - D2 + D3}{A}
\]

where \( D1, D2, \) and \( D3 \) are the areas shown in Figure 2, and \( A \) equals the number of arrivals during Cycle \( k \).

The delay corresponding to area \( D2 \) can be determined from

\[
D2 = \frac{N^2 g}{2C} + N q r
\]

FIGURE 1 Delay model with step arrival flow rate.

FIGURE 2 Delay model with step arrival rate and overflow.
where

\[ N_k = \text{overflow queue at the beginning of Cycle } k, \quad C = \text{capacity per cycle: } C = s g, \text{ where } s \text{ is the saturation flow in veh/sec.} \]

Equation 5 is also used to calculate area \( D3 \) after substituting \( N_k^* \) for \( N_k \).

To determine the area \( D1 \), two cases have to be considered separately: (a) overflow at the end of Cycle \( k \), and (b) no overflow. These cases are shown in Figure 3. Area \( D1 \) can be found as the sum of elementary areas abcd and cdef in Case a or areas abcd and cdg in Case b. After some transformations, the following equation is derived:

\[
D1 = \frac{(2N_k + P_r A) + (A - C) g}{2} + \frac{(C - N_k - A) g}{2[C - (1 - P_r)A]}
\]

when \( N_k + A < C \) and \( C \geq (1 - P_r)A \).

Equations 4–6 determine average delay per vehicle arriving during each signal cycle based on the step arrival flow model. When compared with observations, the estimates can be used to verify the model assumptions. These equations can also be used in macroscopic delay simulation models.

**DELAY OBSERVATIONS**

In order to verify the adequacy of the step arrival flow model, delay was measured at two signalized intersections in Warsaw, Poland, and three intersections in Singapore. Results from
four approaches are presented here. At one intersection (Approach A), the platoons arrived at different times during each cycle because the upstream signal was operating at a different cycle length. At two other intersections (Approaches C and D), the signals were synchronized (same cycle time) but not properly coordinated; as a result, most of the vehicles arrived during the red period. The last intersection studied (E) had a mixed platoon formation and random arrival pattern because part of the flow was coming directly from an expressway.

Traffic flow on all four approaches was filmed for 1 to 3 hr using a video camera. The technique of delay measurement is shown in Figure 4. Arrival and departure times, defined as the times of crossing screenlines A and B, were recorded for all vehicles. Screenline A was established around 200 m upstream from the stop line, beyond the maximum extent of observed queues. Screenline B was located on the other side of the intersection.

The mean observed delay per vehicle was calculated using the following equation:

\[ d_{o_k} = t_{ak} - t_{ak} - T_{1k} \]  

(7)

where

\[ d_{o_k} = \text{average delay incurred by vehicles arriving during Cycle } k, \]

\[ t_{ak} = \text{mean arrival time in Cycle } k, \]

\[ t_{bk} = \text{mean departure time in Cycle } k, \]

and

\[ T_{1k} = \text{average free-flow travel time between points A and B.} \]

The free-flow travel time between points A and B was calculated for each approach from a sample of vehicles that did not stop or were not obstructed by vehicles in front. In addition, the free-flow travel time between point A and the stop line \((T_{2k} \text{ in Figure 4})\) was also determined. Calculating the mean arrival and departure times required identifying the group of vehicles arriving in each cycle. The beginning of Cycle \(k\) was projected back to the screenline A by subtracting free-flow time \(T_{2k}\) from the start of red in that cycle. Thus, the first vehicle arriving during Cycle \(k\) could be identified and used to mark the beginning of a new group. The path of that vehicle was then traced in order to find its departure time. This method ensured that the average times \(t_{a} \text{ and } t_{b}\) were calculated for the same group of vehicles, even though in cases of overflow some of them were departing in subsequent cycles.

The advantage of using this method is that the delay obtained from Equation 7 is the average delay per vehicle arriving during Cycle \(k\) and includes both uniform and overflow delay. Thus, it corresponds to modeled delay given by Equation 4.

Peak-period traffic characteristics obtained from observations at each approach are presented in Table 2. The traffic flow on all approaches could be described as medium to heavy (the degree of saturation ranged from 0.71 to 0.90), with frequent overflow queues observed on Approaches A and D. Table 2 shows the values of the platoon ratio, calculated from Equation 2. Although the average platoon ratio for Approach A was close to 1 (Arrival Type 3), in some cycles up to 80 percent of the vehicles arrived on red, which produced a high standard deviation of delay. Approaches C and E would be

![FIGURE 4 Method of delay measurement.](image-url)
TABLE 2 OBSERVED TRAFFIC CHARACTERISTICS

<table>
<thead>
<tr>
<th>Approach</th>
<th>Mean delay per cycle (sec)</th>
<th>Standard deviation of delay (sec)</th>
<th>Proportion of arrivals during red P_r</th>
<th>Proportion of red in cycle r/t</th>
<th>Platoon ratio R_p</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>25.1</td>
<td>9.7</td>
<td>0.51</td>
<td>0.50</td>
<td>0.98</td>
</tr>
<tr>
<td>C</td>
<td>50.9</td>
<td>4.0</td>
<td>0.81</td>
<td>0.65</td>
<td>0.54</td>
</tr>
<tr>
<td>D</td>
<td>26.3</td>
<td>5.5</td>
<td>0.38</td>
<td>0.29</td>
<td>0.87</td>
</tr>
<tr>
<td>E</td>
<td>39.1</td>
<td>7.7</td>
<td>0.75</td>
<td>0.66</td>
<td>0.74</td>
</tr>
</tbody>
</table>

classified according to their R_p value as having Arrival Type 2, whereas in fact observations indicated that platoons arrived at the beginning of the red period (Type 1).

VERIFICATION OF THE MODEL

To verify the step arrival rate model, delay in each cycle was estimated using Equation 4 and compared with observed values. Table 3 presents the mean of the estimated delay, the deviation between the estimated and observed mean values, root-mean-square error, and the coefficient of correlation. Examples of the comparison are also shown in Figures 5 and 6.

Table 3 and Figures 5 and 6 indicate that the model generally fits the data well (a high coefficient of correlation). The cycle-by-cycle fluctuations of delay at Approach A, which occur because of the changing proportion of arrivals during red, are well reflected by the model (Figure 6a). However, the model tends to underestimate delay for Approaches C and E (the deviation between the estimated and observed means is 5 to 6 sec). This discrepancy was investigated by analyzing the arrival flow profiles during the cycle. Figure 7 shows the mean observed arrival flow rates in 2-sec intervals and the step flow rate model. In both cases, most of the vehicles arrive during the first half of the red period. This process causes additional delay to some vehicles as compared with the step flow rate model, which predicts the mean arrival time during red equal to r/2.

To increase the accuracy of delay estimation, a new model was tested. In this model, a third-degree polynomial function was used to describe the variation of arrival flow rate with time (t). The function had the general form,

\[ q(t) = a_1t + b_1t^2 + c_1 + d \]  

(8)

As in the case of the step model, the only known parameters were assumed to be the mean flow rate (q) and the proportion of arrivals during red (P_r). To determine coefficients \(a_1, b_1, c_1,\)

---

**TABLE 3 COMPARISON BETWEEN OBSERVED AND ESTIMATED DELAY**

<table>
<thead>
<tr>
<th>Approach</th>
<th>Mean estimated delay (sec)</th>
<th>Deviation between estimated and observed delay (sec)</th>
<th>Root-mean-square error (sec)</th>
<th>Coefficient of correlation r</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>24.2</td>
<td>- 0.9</td>
<td>2.8</td>
<td>0.96</td>
</tr>
<tr>
<td>C</td>
<td>44.4</td>
<td>- 6.5</td>
<td>7.1</td>
<td>0.69</td>
</tr>
<tr>
<td>D</td>
<td>25.9</td>
<td>- 0.4</td>
<td>3.5</td>
<td>0.80</td>
</tr>
<tr>
<td>E</td>
<td>33.8</td>
<td>- 5.3</td>
<td>6.3</td>
<td>0.83</td>
</tr>
</tbody>
</table>

---

**FIGURE 5** Comparisons of observed and estimated delay.
FIGURE 6 Cycle-by-cycle comparisons of observed and estimated delay.
FIGURE 7  Examples of arrival flow rate profiles within one cycle.
and \( d \), two additional conditions were imposed: both the value and the slope of \( q(t) \) should be the same at the beginning and the end of the cycle. The four conditions could be written as

1. \( q(0) = q(t_c) \)
2. \( q'(0) = q'(t_c) \)
3. \( \int_0^{t_c} q(t) dt = q(t_c) \)
4. \( \int_0^{t_c} q(t) dt = q(t_c)P \)

From these conditions, the following equation for \( q(t) \) was derived:

\[
q(t) = q \frac{2(P_c + u - 1)}{a^2(1 - u)^2} (2t - 1)(t - 1) + 1
\]  

(9)

where \( t \) equals time expressed as a fraction of cycle time, and \( q \) equals the mean arrival rate.

Curves corresponding to Equation 9 are shown in Figure 7. Generally, they fit the observed arrival flow pattern better than the step flow rate model. Used for computing delay, the polynomial function reduces the discrepancy between estimated and observed mean values by 1.4 and 2.8 sec for Approaches C and E, respectively.

Although the application of Equation 9 seems to improve accuracy, it has the following problems:

1. The variable \( q(t) \) may not be positive. To ensure that the flow rate is positive, allowable ranges of \( u \) and \( P \), should be found.
2. The polynomial curve is not symmetrical inside the red and green periods but has smaller curvature toward the middle of the cycle. This shape helps the accuracy when the arrivals are concentrated toward the beginning of the red period but is counterproductive if the platoon arrives later.

These problems discouraged use of the polynomial function. The step arrival rate model is less accurate when the arrival pattern is not symmetrical inside the red and green periods, but it is generally more flexible.

**COMPARISON OF THE HCM (I) AND STEP ARRIVAL FLOW MODELS**

The HCM (I) and step arrival flow delay models cannot be compared directly because they define delay differently. The HCM (I) uses stopped delay rather than total approach delay. Also, the overflow delay component, which can be found for individual cycles from Equations 3 and 4, in the HCM (I) is defined as an average for a 15-min period. Despite these differences, the relative effect of arrival pattern on delay in both models can be examined.

The arrival pattern is characterized in the HCM (I) method by the value of the platoon ratio. However, the platoon ratio has no clear meaning. Furthermore, according to the step arrival rate model, depending on the signal timing, the same platoon ratio can have different effects on delay. Therefore, the arrival pattern will be described here using the proportion of arrivals both during the red and green time ratios.

The relative effect of platooning on uniform delay can be expressed by the ratio of delay with platooned arrivals to delay with uniform arrivals. According to the step arrival rate model, this ratio for the uniform component of delay will be equal to

\[
f_1 = \frac{d_{up}}{d_u}
\]

(10)

where \( d_{up} \), and \( d_u \) are given by Equations 3 and 1, respectively. Ratio \( f_1 \) defined in this way is equivalent to the progression factor \( p_r \) in the HCM (I) as far as uniform delay is concerned. The values of \( f_1 \) depend on signal timing parameters, degree of saturation, and proportion of arrivals during red.

The relationship between progression factors \( f_1 \) and the proportion of arrivals during red \( P \), is shown in Figure 8. The values were calculated for cycle time \( t_c \) equal to 100 sec, degree of saturation \( (x) \) equal to 0.9, and green time ratio values of 0.3 and 0.6. For comparison, the corresponding HCM (I) progression adjustment factors \( p_r \) are also shown. The step arrival rate model is generally more sensitive to the value of \( P \) than is the HCM (I) model. This difference is especially marked for the larger green time ratio \( (g = 60 \text{ sec}) \), when in the worst case of 100 percent arrivals during red, the delay the step model predicted would be twice that of the HCM (I). At the other extreme, with all arrivals occurring during green, the step model logically predicts zero delay.

Finding the effect of platooning on overflow delay requires a more complicated procedure. First, subtract the uniform delay component \( d_u \) from the total delay given by Equation 4. In this way, the overflow delay component can be calculated for any value of \( P \), including the case in which \( P = 1 - u \) (uniform arrivals). Now, the progression adjustment factor for overflow delay, \( f_2 \), can be calculated as

\[
f_2 = \frac{d_r - d_{up}}{d - d_u}
\]

(11)

where

\[
d_r = \text{total delay with platoon arrivals (Equation 4)};
\]
\[
d_{up} = \text{uniform delay with platoon arrivals (Equation 3)};
\]
\[
d = \text{total delay with uniform arrivals, obtained from Equation 4 with } P_r = 1 - u; \text{ and}
\]
\[
d_u = \text{uniform delay with uniform arrivals (Equation 1)}.
\]

The values of \( f_2 \) depend, among other things, on the overflow queue length \( (N_i \text{ in Equation 5}) \). The relationship between progression factors \( f_2 \) and the proportion of arrivals during red is shown in Figure 9 for several values of \( N \). The other parameters were assumed the same as those used for the curves in Figure 8. The corresponding HCM (I) progression adjustment factors, \( p_r \), are also shown for comparison. Figure 9 shows that the effect of platooning on overflow delay (as predicted by the step arrival rate model) is small. The maximum increase in delay is 6 percent (for \( u = 0.6 \) and \( P > 0.8 \)).

The effect of platooning decreases as saturation increases, and it disappears completely when \( x \approx 1 \). Calculated from Equation 11, \( f_2 \) equals exactly 1 when \( x \approx 1 \) is substituted into the delay equations.
(a) Cycle = 100s, Green = 30s, x = 0.9

Progression adjustment, f1

(b) Cycle = 100s, Green = 60s, x = 0.9

Progression adjustment, f1

FIGURE 8  Effect of proportion of arrival during the red period on uniform delay.
(a) Cycle = 100s, Green = 30s, x = 0.9

Progression adjustment, f2

(b) Cycle = 100s, Green = 60s, x = 0.9

Progression adjustment, f2

FIGURE 9 Effect of proportion of arrivals during the red period on overflow delay.
SUMMARY AND CONCLUSIONS

The HCM (1) method of delay calculation based on arrival types and progression factors is not accurate because the adjustment factors vary in large steps. More accurate models (9,10) exist, but they require additional information such as platoon length and arrival time. These parameters are difficult to measure in practice.

The step arrival rate model is a practical compromise. The model is based on the assumption of different flow rates during the red and green periods in each cycle. The advantage of the step arrival rate model is that it requires only one additional parameter—the proportion of arrivals during the red period ($P_r$), which is not difficult to measure. A delay equation based on this model allows direct computation of the uniform delay component. Delay estimates obtained from the model were found to be in good agreement with cycle-by-cycle delay observations. Discrepancies occurred only when the arrival rate profile was not symmetrical inside the red and green periods. An alternative model, using a polynomial function to describe the arrival rate profile, was found to improve accuracy only in special cases. Therefore, the step arrival rate delay equation is recommended as a general-purpose approximation.

The relative increase or decrease in uniform delay because of platooned arrivals was examined and compared with the HCM (1) progression adjustment factors. Generally, the step arrival rate model predicts a much stronger effect of platooning than the HCM (1) method. This effect is especially visible for larger green time ratios. The demonstrated effect of platooning on overflow delay is not very significant. Therefore, no adjustment factors should be used for the overflow delay component unless a better formula can be found and verified.

REFERENCES


Publication of this paper sponsored by Committee on Traffic Flow Theory and Characteristics.