4.3. Delay Estimation

We now know how to compute the uniform delay, in seconds per vehicle, for a simple approach to a signalized intersection. We have a formula that gives us a valid estimate of the delay to each vehicle under the assumption of uniform arrivals. We must now ask ourselves what happens when the arrivals are not uniform, because we know that there is a random element to the behavior of traffic in the real world. We determined in the last section that the average cycle accommodated 12.5 vehicles. This was based on 500 vph and a cycle length of 90 seconds. We expect, then, that some cycles will accommodate more than 12.5 vehicles while other cycles will accommodate fewer than 12.5 vehicles. The question is “What will this do to the delay estimation?” For that, we have to go back to the queue accumulation polygon.

Let’s consider two successive cycles and say that 15 vehicles arrive on the first cycle and 10 vehicles arrive on the second. This gives an average of 12.5 vehicles per cycle, which is consistent with the example problem that we just solved. To estimate the delay, we must construct a different QAP for each cycle. Because of the different number of arrivals, the two triangles will have different bases and different altitudes, and therefore different areas.

The computations for this example are shown in the box at the right, using the given number of arrivals on each cycle. These values must first be converted to rates per second by dividing by the cycle length of 90 seconds. Knowing the effective red time of 56.7 seconds, the altitude of the QAP triangle may be computed by multiplying the arrival rate by the length of the effective red.

To compute the base of the triangle, we must first determine the saturated green time by dividing the altitude of the triangle (i.e., the maximum queue size) by the net departure rate. We can compute the net departure rate by subtracting the arrival rate from the given saturation flow rate of 0.50 vehicles per sec.

At this point we have both the base and the altitude, so we can compute the area of the triangle to get the total number of vehicle-seconds of delay in each cycle. The total delay for both cycles is simply the sum of the delays for each of the two cycles. The delay per vehicle may then be determined by dividing the total delay by the number of vehicles processed in both cycles, i.e., 25.
Comparing the delay estimated in this manner with the results from the uniform delay formula, we see that the effects of randomness in the arrival rates don’t quite balance out. We naturally accumulate more delay on cycles with above average arrivals, and less delay when the arrivals are below average, but when we add the two delays together, we get a slightly greater total than we would with uniform arrivals.

This is only one of two reasons why the uniform delay equation underestimates the actual delay that you would expect to observe at a signalized intersection. The second reason is the residual queue that accumulates on cycles that are not able to accommodate all of the arrivals. The QAP formulation is based on the assumption that there is no initial queue when the red phase begins. If this assumption is violated, the QAP is no longer a triangle, but a more complex shape with a much larger area. This condition will arise when a cycle fails, by the definition given in Section 6.1. We indicated earlier that the probability of failure can become quite significant as the demand volume approaches the capacity. When the demand exceeds the capacity the QAP area is no longer a valid estimator of delay.

While the terminology, symbols, definitions and computations may differ from model to model, all analytical models of signal operation estimate delay by the following formulation:

\[ D_c = D_u + D_i \]

where:

\[ D_c \] = The control delay, i.e., the delay that would be eliminated if vehicles on the approach were not subject to any form of control.

\[ D_u \] = The uniform delay, determined by computing the area contained within the queue accumulation polygon.

\[ D_i \] = The incremental delay, which represents the additional delay that accrues on the approach because of random and saturation effects in the traffic stream.

The above terminology is used by the 1997 edition of the Highway Capacity Manual (HCM). We will examine the HCM procedures for signalized intersection analysis later. At this point, we will introduce a preview of the HCM formula for the incremental delay.
Let’s begin by plotting the computed delay vs the v/c ratio. We know that the delay per vehicle goes up with the v/c ratio. If we include only the uniform delay term, the relationship is more or less linear as shown in the box on the right.

We also know that the uniform term alone will underestimate the delay and the discrepancy will increase with the v/c ratio. At v/c ratios above 1.0, a perpetually increasing queue will accumulate because, by definition, more vehicles will arrive than the signal can accommodate. Under these conditions, the base of the QAP is the entire length of the period of oversaturated operation, and the altitude is the maximum queue that was attained. The area will therefore depend on how long the demand exceeded the capacity.

The queue accumulation may be calculated easily by a simple deterministic queuing process in which the queue increases at a rate determined directly by the difference between the demand and the capacity. For example, if the capacity is 1000 vph and the demand is 1100 vph, then there will be a residual queue of (1100 - 1000) or 100 vehicles at the end of the period. The residual queue must be serviced during the next period.

For any specified period of oversaturation, the relationship between delay and v/c may be represented by a straight line labeled as the “deterministic queue” in the box at the right. Note that the deterministic queue delay is zero as long as the v/c ratio is less than 1 because there is no accumulation of vehicles over time. When the v/c ratio is greater than 1, the delay associated with deterministic queuing increases very rapidly and quickly goes out of sight.

The control delay (i.e., the sum of the uniform and incremental delays) is illustrated as a separate line in this same graphic. Mathematically, it is represented by a function that is asymptotic to the uniform delay line at low values of v/c and to the deterministic queuing delay at high values of v/c. To achieve this, the incremental delay must be a function that is asymptotic to zero at low values of delay and to the deterministic queuing delay at high values. An example of such a function is:

\[ D_i = 225X^2 \left( X - 1 \right) + \left( X - 1 \right)^2 + \frac{16X}{c} \]

Where

- \( D_i \) = the incremental delay (sec/veh)
- \( X \) = the degree of saturation (v/c ratio)
- \( c \) = the capacity (vehicles per hour)
Note that, in HCM terminology, the cycle length is represented by the upper case C and the capacity is represented by the lower case c. This is sometimes confusing, but it is important to distinguish between these two terms and to use them consistently.

This formula has a built-in period length of 15 minutes. It is very similar to the one that is used in the HCM. We will discuss the differences in a later session.

Now we have all of the tools to estimate the delay on a signalized approach with random arrivals over a wide range of v/c ratios. Let’s apply this formula to our example:

The box at the right indicates that the incremental delay is 4.27 seconds. This is relatively small compared to the uniform delay. The incremental delay term is generally not a significant item until the demand volumes are close to saturation. For example, let’s repeat the full delay computations for v/c = 1.0. This would happen at a volume of 667 vph instead of 500 vph.

When \( X = 1.0 \), the uniform delay term reduces to:

\[
D_u = \frac{0.5C(1-g/C)^2}{(1 - g/C)} = 0.5C(1-g/C) = 0.5(C-g)
\]

In other words, the uniform delay would be one half of the effective red time, or 28.8 sec.

When \( X = 1.0 \) the incremental term reduces to:

\[
D_i = 225 \sqrt{\frac{16}{c}}
\]

which in this case is 34.8 seconds. The control delay will be the sum of the two components, or 63.6 seconds per vehicle. Note that in this case, the incremental term is larger than the uniform term.

We have once more covered a complete set of computations using the northbound approach as an example. As you might expect, the other three approaches will be covered as part of the assignment.

---

**Delay Estimation: NB**

- \( D_u = 25 \) sec
- \( X = 0.75 \)
- \( c = 667 \) vph
- \( 225X^2 = 126.6 \)
- \( X - 1 = -0.25 \)
- \( (X - 1)^2 = 0.0625 \)
- \( 16X/c = 0.018 \)
- \( D_i = 4.27 \) sec
- \( D_c = 29.27 \) sec

**Repeat for v/c = 1.0**

- \( D_u = 28.8 \) sec
- \( D_i = 34.8 \) sec
- \( D_c = 63.6 \) sec

---