Delay at Signalized Intersection

Shivam Gupta
CE 683 Traffic Engineering

Indian Institute of Technology, Kanpur

Abstract

An intersection is a shared space that is used by more than one approach at a time. A signalized intersection is one where the shared space is used alternatively by a fixed number of approaches for a pre defined time interval as per the phasing scheme used for the intersection. Delay and queuing process are the main characteristics of such an intersection which help in analyzing and designing the intersection for a particular level of service.

1 Introduction

Vehicle delay is the most important parameter used by transportation professionals in evaluating the performance of a signalized intersection. This is perhaps because it directly relates to the time loss that a vehicle experiences while crossing an intersection (though we have not considered other problems like congestion due to queuing, extra fuel loss due to vehicle ignition etc.). However delay is a parameter that is not easily determined due to the non deterministic nature of the arrival and departure processes at the intersection. But lot of research has been done in this field to define delay by a number of analytical delay models, including deterministic queuing, steady state stochastic and time dependent stochastic models. There are assumptions in these models that help in simplifying the complex flow conditions to a quantifiable model which gives an approximate measure of average delay faced by a vehicle crossing an intersection. Some research has been done that tries to predict the variance of overall delay that individual vehicles may experience at signalized intersection due to large variation and randomness of traffic arrivals and interruption caused by traffic signal controls.

2 Delay at Signalized Intersection

Delay at signalized intersection is computed as the difference in the departure time and the arrival time of a vehicle. It can also be said equal to the total extra time spent by a vehicle at the intersection than what is required if the vehicle were allowed to pass the intersection without any hindrance. The total delay time can be categorized into deceleration delay, stopped delay and acceleration delay. The deceleration delay is the time loss that the vehicle takes in slowing down to reach to a stoppage, in case the signal is red or to a speed, in case there is a queue which is moving when the signal is green. The stopped delay is the delay that the vehicle spends at the intersection while it is standing in a queue waiting for the signal to turn green. While most of the delay incurred at signalized intersections is directly caused by the traffic signal operation, a fraction of the total delay is due to the time required by individual drivers to react to changes in the signal display at the beginning of the green interval, to mechanical constraints, and to individual driver behavior. When headway of the departing vehicles are observed it is seen that vehicles in front of queue take more time to cross a stationary point than the following vehicles and reach saturation headway after some point. To account for the additional delays due to driver reaction time and vehicle acceleration constraints, the operation of a signalized intersection is usually defined in terms of effective signal intervals instead of actual intervals in delay estimation models. Instead of explicitly considering green, yellow and amber intervals, delay calculations are typically performed by dividing the signal cycle into effective periods of stopped and moving traffic within which constant traffic characteristics can be assumed. The amount of difference between the actual and effective timings will thus depend on the assumptions regarding driver reaction time at the beginning of the green interval and vehicle accelerations.

A final element that may affect the delays incurred at intersection approaches is the randomness in vehicle arrivals. If vehicles were to arrive at uniform intervals, the delays incurred by vehicles within successive signal cycles would be identical, as there would then be an exact replication of the arrival and departure patterns. In this case the flow at one intersection is independent of any other intersection upstream or downstream. In case when the vehicles are arriving in a platoon, intersections may be designed in such a way that the platoon keeps on moving from one intersection to another without getting dissipated. Finally, under random arrival patterns, the number of arrivals may fluctuate from one cycle to the other, thus resulting in different queue lengths. This may in turn result in arrival demands that occasionally exceed the approach capacity, and therefore, in higher delays.

3 Delay Models

3.1 Deterministic Queuing Model

Considering a single lane which is controlled by a traffic signal giving a fraction of time available for movement and assuming that the arrival and departure process are deterministic, this model tries to predict average delay experienced by a vehicle within a signal cycle. First considering unsaturated flow conditions where the vehicles arriving in a cycle are cleared before the start of another cycle. The model is generated by first assuming that the vehicles arrive at a uniform and constant flow rate. This ignores the randomness in arrival and since we are considering only unsaturated conditions it also overlooks on over-saturated flow conditions. A second assumption is that that vehicles decelerate and accelerate instantaneously. This assumption converts all deceleration and acceleration delays into equivalent stopped delay, and thus allows a direct estimation of the total delay incurred by vehicles attempting to cross an intersection. A final assumption is that vehicles queue vertically at the intersection stop line. While this assumption does not represent a normal queuing behavior and may not accurately represent the exact number of queued vehicle at a given instant, it does not bias the delay estimation process over an entire queue formation and dissipation process and is therefore a valid simplification when only considering delay estimations.

$$d_u = \frac{C(1-\lambda)^2}{2(1-X\lambda)} \tag{1}$$

where d_u is average delay due to unsaturated flow, λ is fraction of effective green to cycle length ($\frac{g_e}{C}$), X is approach volume (v) to capacity (c) ratio or saturation ratio (here$X \leq 1$).

In over-saturated conditions, the number of vehicles reaching the intersection exceeds the number of vehicles that can be served by the traffic signal ($X \geq 1$). This causes a growing residual queue to occur. In this case the total delay is categorized into two components, the delay due to unsaturated component and the delay due to over-saturated component.

$$d_o = \frac{C - g_e}{2} + \frac{T}{2}(X-1) \tag{2}$$

where d_o is average delay due to over-saturated flow, T is the time for which the flow is over-saturated.

3.2 Steady State Stochastic Model

While deterministic queuing assumes uniform arrival rate, in practice the arrival process is random and more complex. In order to make it more close to reality, attempts are made to account for randomness of arrivals. First considering that the number of arrivals in a given time interval follows a known distribution, typically a Poisson distribution, and that this distribution does not change over time. Second, they all assume that the headway between departures from the stop line follow a known distribution with a constant mean, or are identical. Third, while it is recognized that temporary oversaturation may occur due to the randomness of arrivals, it is assumed that the system remains unsaturated over the analysis period. Fourth, the system is assumed to have been running long enough to allow it to have settled into a steady state. Fifth, all these models still consider that vehicles decelerate and accelerate instantaneously and thus, that all drivers behave similarly.

Webster's model is one such model. The first term estimates the average approach delay assuming uniform arrivals. The second term considers the additional delays attributed to the randomness of vehicle arrivals. The third term is an empirical correction factors that reduces the estimated delay by 5-15%, to be consistent with simulation results.

$$d = \frac{C(1-\lambda)^2}{2(1-X\lambda)} + \frac{X^2}{2v(1-X)} - 0.65\left(\frac{c}{v^2}\right)^{1/3} X^{2+\lambda} \tag{3}$$

3.3 Time Dependent stochastic Model

A main consequence of the stochastic delay modelling described in the previous section is that the estimated delays tend to infinity as traffic demand approaches saturation ($X \geq 1$). In order to explain more rigorously the delay for all flow conditions, a general time dependent delay model is conceived primarily based on empirical calculations. Considering the Highway Capacity Manual (2000) model, the average delay can be given by

$$d = d_1 \cdot f_{pf} + d_2 \tag{4}$$

$$d_1 = 0.5C\frac{(1-\lambda)^2}{1-\lambda \cdot x_1}$$

$$d_2 = 900T\left((X-1) + \sqrt{(X-1)^2 + \frac{8kX}{cT}}\right)$$

where f_{pf} is delay adjustment factor for quality of progression and control type (= 1.0 for pre-timed non coordinated signals), c is the capacity of lane (in vph) and k is an incremental delay factor dependent upon signal controller setting (0.50 for pre timed signals), T is the time of evaluations (in hr) and x_1 is $min(1.0, X)$. This model assumes steady-state traffic conditions. It estimates delay under stochastic equilibrium conditions, when the arrival and departure flow rates have been stationary for an indefinite period of time. It also assumes that the number of arrivals in a given interval follow a Poisson distribution and that the headway between departures have a known distribution with a constant mean value. Similar to the earlier models, it also assumes instantaneous acceleration and deceleration to simply data estimation

4 Variability in Delay

The models that have been described give average delay. Since delay is an important parameter that is used as a decision variable in designing the intersection it is desired that a distribution is predicted that would approximate the delay and match the average delay proposed by these analytical models. Consider the cumulative arrival and departure of vehicles during the time interval [0,T] The delay for a particular vehicle arriving at time t, called *overall delay* and noted as D, is considered to include two components: *uniform delay* (D_1), when the approach is un-saturated and vehicles arrive uniformly and *overflow delay* (D_2), due to temporary overflow queues resulting from random nature of arrivals. Similarly the total variance of delay can be divided into two components, $var[D_1]$ and $var[D_2]$.

$Var[D_1]$, represents the variation of uniform delay that would be experienced by vehicles arriving during time interval [0,T]. This variation results from the uncertainty of the vehicle's arrival time during each cycle of the interval. The vehicle can arrive at any moment within a cycle and thus experience variable delays as a result of the signal control. An estimate of this variance component can be obtained theoretically on the basis of a deterministic queuing model with vehicles arriving uniformly during the cycle

$$\text{var}[D_1] = \frac{C^2(1-\lambda)^3(1+3\lambda-4\lambda x_1)}{12(1-\lambda x_1)^2} \tag{5}$$

In order to establish a model for the variance of delay caused by an overflow queue, two extreme traffic conditions are first investigated: unsaturated conditions ($X \leq 1.0$) and over-saturated conditions ($X > 1.0$). For unsaturated conditions, overflow delay experienced by a vehicle arriving during the time interval [0,T] is mainly caused by occasional overflows of traffic from each cycle. The relationship between the variance of this delay and the degree of saturation can be approximated by

$$\mathrm{var}[D_2] = \frac{X(4-X)}{12c^2(1-X)^2} \tag{6}$$

For $X = 1$, the variance becomes infinite which is incorrect. This expression gives good approximate for the variance for values of $X < 1$. If the intersection approach is highly over-saturated during time period [0, T], there is a high probability that an overflow queue always exists during the period from time 0 to time T. Consider a vehicle arriving at time t during time period [0,T]. The overflow queue for a vehicle arriving at time t, Q_t , can be determined as the total arrivals minus the total departures and then the delay D_2 will then be given by the time taken for the queue to dissipate with capacity flow rate.

$$Q_t = N_t - c \cdot t$$

$$D_2 = \frac{N_t - c \cdot t}{c}$$

where N_t is a random variable with mean vt. The variance of delay D_2 for vehicles arriving during time interval [0, T] can be obtained by assuming that the arrival time t is a random variable with known distribution. The term D_2 becomes random and which is dependent upon time. Therefore using the total law of variance and law of conditional probability, variance of D_2 can be written as (for detailed description please refer Appendix (B)

$$\begin{aligned}
\mathrm{var}[D_2] &= \frac{E[\mathrm{var}[N_t - ct|t]] + \mathrm{var}[E[N_t - ct|t]]}{c^2} \\
&= \frac{E[\mathrm{var}[N_t|t] - 0] + \mathrm{var}[E[N_t|t] - ct]}{c^2} \\
&= \frac{E[\mathrm{var}[N_t|t]] + \mathrm{var}[vt - ct]}{c^2} \\
&= \frac{E[\mathrm{var}[N_t|t]] + (v-c)^2\mathrm{var}[t]}{c^2}
\end{aligned}$$

Assuming that the arrival time is uniformly distributed during the time interval [0,T] with mean equal to $T/2$ and variance equal to $T^2/12$, and since the arrival follows Poisson distribution (for which mean is equal to variance), using $\mathrm{var}[N_t|t] = E[N_t|t]$:

$$\mathrm{var}[D_2] = \frac{TX}{2c} + \frac{T^2(1-X)^2}{12} \tag{7}$$

It must be emphasized that Eq.(7) is valid only when an overflow queue is present during the period from time 0 to time t. In reality, however, it is possible that no overflow queue exists at time t, and consequently no overflow delay is experienced. Therefore, it can be concluded that Eq.(7) represents an upper-bound estimate of the variance of overflow delay. The actual variance would be lower, but the prediction error should become smaller as the degree of saturation increases and the associated likelihood of overflow queuing increases. This can be done by multiplying Eq.(7) with a correction factor $e^{-\left(\frac{x_0}{X}\right)^{\beta}}$ where values of parameter x_0 and β are calibrated using simulated data of variances.

References

[1] L. Fu, and B. Hellinga, Delay Variability at Signalized Intersections. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1710*, TRB, National Research Council, Washington, D.C., 2000, pp. 215-221.

[2] F. Dion, H. Rakha, and Y. S. Kang, Comparison of delay estimates at under-saturated and over-saturated pre-timed signalized intersections, *Transportation Research Part B (38)* (2004), pp. 99-122.

Appendices

A Law of Total Expectation

Assuming that the events X and Y are random events in the same probability space, then the expected value of conditional expected value of X given Y is the same as the expected value of X. This is proved as:

$$\begin{aligned}
E\left[E[X|Y]\right] &= \sum_y E[X|Y=y] \cdot P(Y=y) \\
&= \sum_y \left(\sum_x x \cdot P(X=x|Y=y)\right) \cdot P(Y=y) \\
&= \sum_y \sum_x x \cdot P(X=x|Y=y) \cdot P(Y=y) \\
&= \sum_x \sum_y x \cdot P(Y=y|X=x) \cdot P(X=x) \\
&= \sum_x x \cdot P(X=x) \cdot \left(\sum_y P(Y=y|X=x)\right) \\
&= \sum_x x \cdot P(X=x) \\
&= E[x].
\end{aligned}$$

B Law of Total Variance

Assuming that X and Y are random variables in the same probability space and using the law of total expectation given in Appendix A, the total variance of X can be written as:

$$\begin{aligned}
\mathrm{var}[X] &= E[X^2] - E[X]^2 \\
&= E\left[E[X^2|Y]\right] - E\left[E[X|Y]\right]^2 \\
&= E\left[\mathrm{var}[X|Y] + E[X|Y]^2\right] - E\left[E[X|Y]\right]^2 \\
&= E\left[\mathrm{var}[X|Y]\right] + \left(E\left[E[X|Y]^2\right] - E\left[E[X|Y]\right]^2\right) \\
&= E\left[\mathrm{var}[X|Y]\right] + \mathrm{var}\left[E[X|Y]\right]
\end{aligned}$$