Delay Estimation at Signalized Intersections with Variable Queue Discharge Rate

Mohsin Sahzad CHAUDHRY\textsuperscript{a}, Prakash RANJITKAR\textsuperscript{b}

\textsuperscript{a,b}Department of Civil and Environmental Engineering, University of Auckland, Auckland 1142, NEW ZEALAND
\textsuperscript{a}Email: mcha341@aucklanduni.ac.nz  
\textsuperscript{b}Email: p.ranjitkar@auckland.ac.nz

Abstract: The techniques developed for delay estimation in the most traffic signal design guidelines are based on the assumption that queue discharge rate at signalized intersections becomes stable after a few vehicles passes through the stop line, which is termed as saturation flow rate. This assumption has been challenged in recent times as a number of field observation in different parts of the world reported an increasing queue discharge rate observed along the back of queue. This paper proposes an empirical model that is capable of capturing the queue discharge behavior observed at signalized intersection. The model is implemented to estimate delay and compared with the existing delay models. The results revealed that the proposed model can overcome the deficiencies of the existing models and can estimate delay more accurately.

Keywords: delay model, HCM 2010, ARR 123

1. INTRODUCTION

Delay is an important measure widely used in traffic engineering to evaluate the operational efficiency of signalized intersections. It is one of the main factors on which level of service rating is determined. Beckmann et al. (1956) were among the pioneers who studied and presented delay formulas for signalized intersections based on queuing theory. Later, Webster (1958) extended his work to estimate signal timing and delay at a signalized intersections based on a simplified assumption that when signal turns to green the queue discharge rate across the stop line increases rapidly until it reaches at a sustained maximum level termed as saturation flow rate where it remains stable (unchanged) until either the queue is exhausted or the green phase ends. This assumption makes it considerably easy to calculate the lane group capacity, which is equal to a product of saturation flow rate and green to cycle time ratio (g/C).

Several factors were identified in the past influencing the saturation flow rate; however the concept remained the same as proposed by Webster (1958). Figure 1 presents this traditional model of traffic signal design at signalized intersections, which is based on the assumption that the saturation flow rate remains constant for a fully saturated intersection for all portions of the green interval except at the beginning and at the end.

The concept of saturation flow rate is the basic parameter in estimation of almost all performance indicators at signalized intersections including delay. Delay is the single most important factor that drivers can perceive and there it drew a lot of attention of past researchers. Webster (1958) was among those who presented a delay formula expressed as follows;

\[
d = \frac{C(1-\lambda)^2}{2(1-\lambda x)} + \frac{x^2}{2q(1-x)} - 0.65 \left(\frac{C}{q^2}\right)^{1/3} \times x^{(2+5\lambda)}
\]
where,

- $d$ is the average delay per vehicle in seconds
- $C$ is the cycle time in seconds
- $\lambda$ is the portion of the cycle which is effectively green for the phase under consideration i.e. $g/C$.
- $x$ is the degree of saturation.
- $q$ is the arrival rate in vehicles per second.

The delay model in Eq (1) comprised of three main elements. The first two terms were derived based on queuing theory where the first term estimates delay assuming the traffic arrivals and departures are completely uniform and the second term assumes a steady state condition and accounts for randomness in arrival rate that leads to overflow queues. The third term was proposed as a correction term based on empirical observations to compensate overestimation of delay when using only the first two terms.

While a great deal of investigations was conducted in the past to quantify the effects of varying arrival flow rate on delay at signalized intersections, the variability in departure flow rate could not get much attention. The assumption of a constant discharge flow rate was one of the reasons for this low attention. A number of studies conducted on queue discharge behavior at signalized intersection reported to have observed large variations in the saturation flow rate (Chaudhry and Ranjitkar, 2013, Chaudhry et al., 2011, Li and Prevedouros, 2002, Lin et al., 2007, Lin and Thomas, 2005, Lin et al., 2004, Tarko and Tracz, 2000, Teply, 1983). Teply (1983) noted that the saturation flow rate depends not only on site-specific conditions but also on the duration of green period and type of community. In a study conducted in Canada, he observed that the maximum queue discharge rate usually drops after about 50s of green interval. A similar study conducted in Hawaii, USA by Li and Prevedouros (2002) reported a rather complex relationship between the saturation flow rate and queue position. It was noted that the minimum headway was not reached until the 9th to 12th vehicle crossed the stop line in queue for through and left turn movements.

A series of investigations conducted more recently in Taiwan and USA revealed a more consistent increase trend in the saturation flow rate (Lin et al., 2007, Lin and Thomas, 2005, Lin et al., 2004). Lin et al. (2004) noted that in Taiwan queue discharge often does not confirm to the notion to a quick rise to a steady state. Similar trends were observed at three...
intersections in Long Island, New York (2005). They all exhibit a general trend of gradual compression of headways as the queue discharge continues. Lin et al. (2007) quantified the extent of errors in the observed data based on statistical analysis conducted on 38 urban lanes in Taiwan. He noted that the discharge rates were increased on average by 24% for straight through movement and 16% for protected left turn movement when compared with HCM approach. They also noted that there is a 40% chance that lost time will differ from the correct value by 2 sec and 50% chance that the estimated capacity will deviate from the actual capacity by 5%, if average HCM lost time and saturation flow of a group of similar lanes are used as estimates for each lane in the study group.

The strong evidence of variations in queue discharge flow rate reported in the literatures suggests a need for further investigation to verify such trends and then incorporate it in the delay formulation, which might contribute to improve the accuracy of delay estimation. This paper proposes an empirical model incorporating an increasing queue discharge rate observed at six signalized intersections in Auckland, New Zealand. The model is implemented to estimate delay at signalized intersections and then compared with the existing methods of delay estimation. Delay models proposed in literatures are reviewed in the following section; followed by results from field observations in Auckland, New Zealand. An Empirical model is proposed that incorporate an increasing queue discharge rate. The impact of this model on the delay is investigated in subsequent section. Finally, some concluding remarks are drawn in the last section.

2. DELAY MODELS

In signalized intersections, delay is the difference between the actual travel time a vehicle experience in traffic control and the travel time in the absence of any traffic signal control. The calculation of the delay depends on several factors including probabilistic distribution of arrival flow (demand), signal timings and departure flow rate (supply), and the time when the vehicle arrives at the approach. Many of these factors are highly variable, thus making estimation of delay very complicated. The level of service (LOS) criteria for signalized intersections is set on estimation of delay. The LOS is an indicator of operational efficiency of the intersections by which quality of service is determined. The value of LOS is represented by letters A through F, with A being best and F being worst. Delay is used to determine the level of service of a signalized intersection, being the only element that is truly perceived by the drivers.

Delay experienced at traffic signals can be divided in to two components: uniform delay and incremental delay. Uniform delay can be estimated using deterministic queuing approach considering a simple case of D/D/1. It is a well-known fact that during the congested period, the arrival flow rate is approaching to a uniform state. In traditional approach, assuming a uniform rate for arrivals and departures makes it a simple case of area calculation to estimate delay. For this purpose, an assumption is made that all vehicles accumulated during phase passes during green time. Solving this case results in formulation of the first term of Webster equation of delay Eq (1). This assumption cannot be implemented on isolated signalized intersections where the flow pattern is randomly distributed. To resolve this issue, a component of random delay equation is introduced that assume Poisson distribution for arrivals (Kendall, 1951). The random delay or overflow delay component includes the portion of delay that occurs due to temporary overflow of queues resulting from the randomness in the arrival rate. The random delay is an additional term introduced to incorporate delay component above uniform delay. This random delay component is adopted as a second term in the Webster delay equation.
Theoretically, incorporation of these two terms should represent actual delay, however it was observed that the first two terms produces a higher value of delay. Therefore a third term was introduced which was empirical in nature and it was derived from the simulation of traffic flow and generally refers to correction term (Webster, 1958). Webster formula for delay was later refined to eliminate correction term and a factor is introduced instead to reduce the sum of first and second term by 10%. (Courage and Papapanou, 1977)

One of the major drawbacks of the Webster’s model was its inability to compute delay at saturation level (x≈1). Webster’s model performs reasonably well in under-saturated (x<1) condition. However when the approach frequently faces a condition during which accumulated queue cannot dissipate fully in one cycle, a phenomenon of growing queue is developed, which is termed as overflow delay or incremental delay as referred in HCM (2010). Akcelik (1981) developed a formula to overcome this shortcoming of Webster model for overflow delay component;

\[ d_2 = \frac{cT}{4} \left[ (x - 1) + \sqrt{(x - 1)^2 + \left( \frac{12(x - x_0)}{cT} \right)} \right] \]  

where \( T \) is analysis period duration (h) and \( c \) is capacity (veh/h) and other variables are as previously defined.

Later, this model in Eq (2) was incorporated in HCM with some modifications. The HCM model for signalized intersection contains three terms;

\[ d = d_1PF + d_2 + d_3 \]  

where,

- \( d \) is average signal delay per vehicle in seconds
- \( d_1 \) is average delay per vehicle due to uniform arrivals in seconds
- \( PF \) is progression adjustment factor
- \( d_2 \) is average delay per vehicle due to random arrivals in seconds
- \( d_3 \) is average delay per vehicle due to initial queue at start of analysis time period, in seconds

The average delay due to uniform arrivals is computed with the following equation:
\[ d_1 = \frac{0.5C(1 - \frac{g_c}{c})^2}{1 - \min(1, x) \cdot \frac{g_c}{c}} \] (4)

The incremental delay formulation in HCM 2010 is as follows:

\[ d_2 = 900T \left[ (x - 1) + \sqrt{(x - 1)^2 + \left( (x - 1)^2 + \frac{4x^2}{cT} \right)} \right] \] (5)

The strong evidence of variations in queue discharge flow rate reported in the literatures suggests a need for further investigation to verify such trends and then incorporate it in the delay formulation, in order to obtain better accuracy in delay estimations.

3. FIELD OBSERVATIONS IN AUCKLAND

To verify the variability in saturation flow rate, a study is conducted based on data collected from video recording on six intersections in Auckland, New Zealand. A number of site selection criteria parameters were established in order to locate the ideal intersection without applying adjustment factors for prevailing conditions in the light of recommendations made by Le et al. (2000). Three main selection criteria includes presence of heavy traffic with at least one exclusive through lane, the ideal geometric and roadway conditions with at least 3.6m lane width, level approach grade, minimal or no pedestrian movements, no curbing parking, no bus stop in the vicinity and acceptable distance from adjacent intersection. The data was collected during peak hours with minimum 2 hours of data recording on all intersections except one intersection in fair weather conditions. The sites selected include Dominion – Balmoral Road Intersection, Balmoral – Sandringham Road Intersection, Great South Road – South Eastern Highway, St Lukes – New North Road, Manukau – Greenlane East, and Pah – Mt Albert Road.

Individual headways are recorded for each vehicle in queue. The data analysis process involved collecting the headways between successive vehicles for two hour time period for each of the six intersections. Initial examination of the headways for each intersection shows some differences in the driver’s reaction time as shown in Table 1. Five intersections are located in the close proximity and show the reaction time varying from 1 second to 1.2 seconds. The results of initial examination of Great South Road – South Eastern Highway intersection shows a different trend than the remaining five intersections with average Reaction Time of 2.05 seconds. The start-up lost time observed within the range of 1.18 seconds to 3.08 seconds. The high start-up delay at St Lukes – New North Road is probably due to downstream approach grade with a sharp curve that resulted in increases of start-up lost time.

The possible reasoning of this increasing trend could be the driver’s behavior towards the end of queue to pass the intersection before signal changes to red. This driver behavior indicates the need of the study to look into the car following variables at signalized intersections and further investigations are required which is out of scope for this study. The results are summarized in the Table 1. These results give a clear indication that there is a relationship between queue discharge rate and the green time as previously reported in some studies. Based on these results, an empirical model is proposed to predict the varying nature of queue discharge flow rate.
Table 1: Results of studies and relationship between green time and queue discharge rate

<table>
<thead>
<tr>
<th>Intersection</th>
<th>Number of Phases</th>
<th>Reaction Time (Start of green to movement of first vehicle)</th>
<th>Start-up Lost Time</th>
<th>Queue Discharge Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balmoral - Sandringham Rd</td>
<td>58</td>
<td>t = 0.79</td>
<td>R = 1.72</td>
<td>Equation: 0.0099 t + 0.3496</td>
</tr>
<tr>
<td>Balmoral - Dominion Rd</td>
<td>59</td>
<td>t = 0.57</td>
<td>R = 1.26</td>
<td>Equation: 0.004 t + 0.4853</td>
</tr>
<tr>
<td>GT South Rd - SE Highway</td>
<td>22</td>
<td>2.05 t = 0.57</td>
<td>R = 1.22</td>
<td>Equation: 0.0098 t + 0.4113</td>
</tr>
<tr>
<td>Manukau - Greelane East Road</td>
<td>58</td>
<td>1.15 t = 0.45</td>
<td>R = 1.12</td>
<td>Equation: 0.0027 t + 0.4711</td>
</tr>
<tr>
<td>Pah - Mt Albert Road</td>
<td>60</td>
<td>1.18 t = 0.41</td>
<td>R = 1.41</td>
<td>Equation: 0.0045 t + 0.4618</td>
</tr>
<tr>
<td>St Lukes - New North Road</td>
<td>54</td>
<td>1.00 t = 0.47</td>
<td>R = 0.96</td>
<td>Equation: 0.0098 t + 0.3842</td>
</tr>
</tbody>
</table>

Note: t is the time passed after onset of green

4. MODEL FORMULATION

Based on the field observations, an empirical model is proposed to accommodate expected increasing queue discharge rate behavior at signalized intersections. The details of the empirical model have been presented in another paper (Chaudhry and Ranjitkar, 2013). The model proposed is nonlinear in nature and is in the following shape;

$$ Q = \frac{t - l_s}{1 + h_m (t-\infty)} $$

where $Q$ is discharge rate in veh/sec, $t$ is the time in seconds, $l_s$ is the initial lost time in seconds, $h_m$ is minimum average headway recorded after a number of cycles and $\infty$ is a correction factor. Here $Q$ is discharge rate after $t$ seconds of the movement. The analogy of the maximum discharge rate is not used here, as the maximum discharge rate is highly variable at the initial stages, instead a discharge rate achievable in time $t$ is used. This discharge rate is dependent on the lost time a signal management parameter encounter during the start of the green time. This model also incorporates the initial reaction time of the drivers after seeing the green time and start responding. This reaction time and response time varies with the type of transmission of vehicles and drivers’ behavior. $h_m$ is the minimum average headway recorded after a number of cycles.

4.1 Derivation of Delay under Uniform Arrivals

A D/D/1 case is considered to derive a delay formula for uniform arrivals and varying departure rate as shown in Figure 3. In order to coincide with the existing methodologies being used in practice, certain assumptions are necessary to make. The first assumption made in this derivation is pertaining to the signal capacity which is exceeding the arrival flow rate. The second assumption is that there is no initial queue at the start of the green time. Lastly it is assumed that the queue formed during the red phase dissipates during the green time in the same cycle. These assumptions make it possible to establish a point where arrival line and departure line meet after vehicles are allowed to move during green phase.

The arrival flow rate line gives the total number of vehicle arrival at time $t$ as shown in Figure 3. A red phase breaks the flow at signalized intersection for $r$ seconds, and then at onset of green, the flow continuous with the departure flow rate. The dotted line indicates the traditional discharge flow rate and broken line indicate the modified model discharge flow rate. Under the traditional concept of uniform arrival and departure rate, the problem of delay can be solved as presented in the first term of Webster Eq. (1). The derivation of the first term
can be referred to any textbook on traffic engineering. When departure rate is varying, a modified form of Eq. (6) can be used in D/D/1 queuing analysis, replacing \( g \) with \( t \) and assuming \( \alpha \) and \( l_s \) as zero;

\[
\frac{t}{1 + h_m(t)}
\]

The assumption of more capacity than arrivals means that all the vehicles that come in a cycle are cleared within the same cycle as shown in Figure 4. In absence of constant departures, a new line is formed that represent a nonlinear model as proposed in Eq (7). In Figure 3, at horizontal axis, point o-r is denoting the red time \( r \), \( r-g \) and \( r-g' \) representing the time required to completely dissipate the queue \( t_D \). Multiplying departure flow rate with time \( t_D \) gives the departures

\[
\text{Number of vehicles departing} = \frac{t_D^2}{1 + h_m(t_D)}
\]

The slope of cumulative arrival line represents the uniform arrival rate approaching at the signalized intersection. The red time stops the traffic flow during which departure flow is zero. During the green time, the slope of cumulative departure line remains zero and equal to \( s \) at the onset of green signal for that approach.

\[
\text{Number of vehicles arriving in time}(r + t_D) = v(r + t_D)
\]

The intersection of the arrivals and departures can be calculated by equating arrivals and departure terms.

\[
v(r + t_D) = \frac{t_D^2}{1 + h_m(t_D)}
\]

Simplifying above equation by using quadratic equation and ignoring negative value, the time required to dissipate the queue can be calculated as;

\[
t_D = \frac{(v + v_h.m.r) + \sqrt{(v + v_h.m.r)^2 + 4vr}}{2(1 - v_h.m)}
\]

Integrating the arrival triangle and deducting the area under the curve of departures gives this expression;

\[
D_t = \frac{v}{2} (t_D + r)^2 - \left[ \frac{t_D(h_m t_D - 2)}{2 h_m^2} + \frac{log(h_m t_D + 1)}{h_m^3} \right]
\]

where \( D_t \) is Aggregate uniform delay, and \( v \) is traffic flow, \( r \) is length of red phase and \( s \) is saturation flow rate.

The effect of log term is related to the gain of discharge rate at onset of green time, which becomes insignificant in this case, so neglecting it.

\[
D_t = \frac{v}{2} (t_D + r)^2 - \frac{t_D(h_m t_D - 2)}{2 h_m^2}
\]
Eq (11) and Eq (12) estimate aggregate uniform delay in vehicle – seconds for one signal cycle. To get an estimate of average uniform delay per vehicle, the aggregate delay is divided by the number of vehicles arriving during the cycle,

\[ \bar{d} = \frac{D}{N} \]  

(14)

where \( \bar{d} \) is average uniform delay per vehicle and \( N \) is number of arrivals during cycle time \( T \) and \( N \) can be represented as;

\[ N = vC \]  

(15)

**Figure 3:** Signalized intersection queuing with traditional and modified model II

### 4.2 Incremental Delay

The component of incremental delay used in HCM 2010 is also incorporated in the proposed nonlinear model. The model proposed in HCM 2010 is in form of a general time-dependent delay model which was conceived in late seventies (Kimber and Hollis, 1979, Robertson, 1979). Empirical evidence indicates that this model predict reasonable results, though no rigorous theoretical basis for this approach is reported (Dion et al., 2004). A coordinate transformation technique is used to transforms the equation that defines a steady-state stochastic delay model to produce asymptotic to the deterministic over-saturation model. Due to pure empirical nature of this equation, not a direct derivation is made and instead same model is used after replacing volume to capacity ratio from modified model.

### 5. MODEL VALIDATION

In order to verify the proposed uniform component of delay, an example case is considered in which an arrival flow rate of 1200 vph is analyzed with the traditional model and with modified model. 30 seconds of red time break the traffic flow pattern, and after onset of green, traditional model predicted a 60 seconds time to a state of arrival flow pattern. The modified model indicated that after 53.32 seconds, the traffic flow will come back to arrival...
flow pattern. This shows the saving of 6.68 seconds of green time ($g' - g$) which is about 11% that can be allocated to next phase as shown in Figure 3.

In order to verify and test the performance of the proposed model, two cases were analyzed. In one case, the incremental delay model of HCM2010 is used without making any changes. In second case, the v/c ratio and average capacity c is replaced with proposed nonlinear model in incremental delay component of HCM 2010. The first term in both cases were replaced by the proposed model. The results indicated that the calculated delay values are closely lying with the delay model of HCM as shown in Figure 4. In first case, when uniform term is replaced only, the results shows a lower delay at degree of saturation between 0.85 to 0.95 and then a gradual increase surpassing HCM 2010 delay value in over-saturation state. The curve formed due to both terms replacement indicates a further compression at v/c ratio of 0.8 to 1.0. The curve then gradually surpass HCM 2010 curve but remained below the curve of first case.

![Figure 4: Delay estimate comparison between HCM, Webster and Model II](image)

### 6. DISCUSSION

From the early days when delay models for signalized intersections are proposed, it was observed that the delay predicted by the model is about 5 to 15% exceeding from the actual delay. Webster (1958) realized it first and introduced a correction term. The approximate over-estimation of delay goes as high as 15% of the actual delay; however an average value remains 10% of the sum of the uniform and the random delay component (Roess et al., 2010). The first term of traditional delay models is based on a uniform pattern of arrivals and departures. Adoption of new model based on the variable departure flow rate provides some of the reasoning behind this over-estimation in delay from traditional approach. The results indicated that about 7% decrease in delay is recorded after 45 seconds of green time from the traditional model. A comparison of the 1st term of Webster model and proposed non-linear model is shown in Figure 5. The lower volume of at a particular approach is indicative of a lower green time. At the lower green time, the non-linear model predicts a higher delay value which is obvious because of the reason that the delay curve of Webster model is based on an
average value of Saturation flow rate. For lower green times, the queue discharge rate observed is lower and therefore the non-linear model predicts a higher delay than Webster model. Later when the queue discharge rate improves, gradually departing curves appear where non-linear model predicts lower delay than Webster model.

Figure 5: Delay Comparison between Webster and non-linear model for Uniform Delay

Figure 5 indicates another benefit of adopting variable departure flow rate. Incorporating variability in the departure flow rate can help in predicting relatively accurate performance measures during the short green times. Traditional approach cannot make distinction between short cycle times and long cycle times which frequently occur on signalized intersections. Although the impact of short green time is not significant, however vehicles may have to wait for next green cycle due to the reason green times are allocated on the basis of fixed saturation flow rate.

7. CONCLUDING REMARKS

This paper has implemented an empirical model to estimate delay at signalized intersection. The variations at departure flow rate can impact the delay calculations. A decrease in delay estimation reduces the need for a correction term that was recommended by Webster (1958) due to over estimation of delay. The reduction in delay in first term indicates that the queue is dissipated earlier during long green cycles and the saving of the green time gained from this early dissipation can be utilized to other phases. For a single cycle time, the effect of this reduction in delay might not be significant however on a large network where a number of intersections can add up the delays, this overall saving is significant.

The empirical model proposed in this paper is capable of capturing the queue discharge behavior observed at six signalized intersections in Auckland. The proposed delay model is compared with the existing delay models. The results show that the proposed model can better approximate delay for uniform arrival rate. This investigation confirms the findings of the previous researchers that Webster model overestimates the delay. The incorporation of variable discharge flow rate in the uniform component of the delay formulation lowered the delay estimation by 5 to 6%. This decrease in delay compensate for a significant proportion of delay overestimation by Webster formulation which is approximately 10%. The results revealed that the proposed model can overcome the deficiencies of the existing models and can estimate delay more accurately.
REFERENCES


310-321.
Webster, F. V. (1958) *Traffic signal settings*, HMSO.