TIME DEPENDENT DELAY AT UNSIGNALIZED INTERSECTIONS

Werner Brilon, Ruhr-University Bochum
Universitaetsstrasse 150, D-44780 Bochum, Germany
E-mail: werner.brilon@rub.de

ABSTRACT

Calculation of intersection delay is usually based on methods obtained from queuing theory. Due to the variability of traffic demand over time, the estimation of delays for time-dependent flow and capacity, where also a temporary overload is allowed, is of primary interest. For solving this problem, a variety of methods is used in current practice. All of these solutions are only approximations. One first step of approximation is the assumption that the priority system can be modeled by an M/M/1-queue. The second step is the so-called coordinate transformation technique. For this method, three sub-groups can be defined. The paper investigates the background of the possible solutions and the quality of approximation. As a basis, a classification of potential delay formulas is defined. This classification accounts for the kind of sophistication of the approximation and for the kind of delay definition, which is treated as the average. Over all, nine useful classes of formulas can be defined. For each of these classes, delay formulas are derived. Some of them correspond to well-known results. However, in addition to that, the complete set of results offers new solutions - also for more realistic cases. Thus, also initial queues at the beginning of the observed peak period as well as different conditions in the post-peak period can be described. As methods for validating these formulas, a Markov-chain formulation has been developed to produce numerically exact results. Also stochastic simulations and empirical data are used for comparison to check the approximate solutions against reality. As a result, a set of equations is available which can be applied to estimate average delays at unsignalized intersections for well-defined traffic conditions. The paper makes clear that instead of an uncritical use of delay formulas, a well sophisticated selection of the adequate equation is also required in practice.
INTRODUCTION

For unsignalized intersections the average delay has been chosen as the representative measure of effectiveness for characterizing the performance of traffic flow by most of the guidelines used in practice (USA: HCM 2000, chapter 17; HBS 2001, chapter 7). Therefore a reliable and realistic estimation of average delay is of significant importance in traffic engineering. Guidelines use several different methods for calculating delays; e.g. the HCM uses the Akcelik, Troutbeck (1991) method whereas the German HBS recommends the rather complex formula proposed by Kimber, Hollis (1979; eq. 22). Both lead to different results. There is no clear insight, which of both is more correct or if alternative solutions might be even better.

The important aspect of the average delay estimation at intersections is the treatment of a temporary oversaturation. Here delay estimation turns out to be one specific application of a more general problem of mathematical queuing theory. The unfortunate fact is that useful analytical solutions are not offered from mathematical theory. Therefore engineering scientists have developed approximate formulas (Kimber, Hollis 1979; Akcelik, Troutbeck 1991 and other publications by Akcelik; Brilon, 1995).

Most of these approaches use the derivation of time dependent patterns for queue length as a starting point. A rather precise but still approximate method to estimate time-dependent queue length distributions are the differential equations given by Newell (1982; see also Troutbeck, Blogg 1998) derived from his so-called diffusion theory. Such time-dependent queue length functions can be integrated to derive delay parameters like average delay. Such an approach seems, however, to be too complicated for practical application. Moreover, if all solutions are still of approximate nature then for the sake of easy application it seems to be desirable to find solutions directly on the scale of delays.

This paper shows that the well-known solutions for average delays at unsignalized intersections are specific members of a larger family of approximate solutions, where other elements of the whole set of delay equations provide more advantages. A classification of different possibilities for average delay definition is also offered. Among all possible cases the traditional formulas only constitute a solution for rather specific and unrealistic cases. These considerations also make clear that a sound understanding of the formulas’ background is crucial for a correct application. The key to the assessment of a useful delay estimation technique is, however, the comparison with exact results, which for this paper have been elaborated using Markov chain techniques, simulations, and some empirical data.

THE BASIC MODEL

Our considerations are concentrated on the simple case of one major stream and one minor stream where the vehicles from the minor stream have to give priority to major street vehicles (= “priority system”; cf. Figure 1). This priority system can be modeled by a queuing system where the first space for a vehicle next to the stop line is treated as the service counter whereas the further waiting positions for minor street vehicles form the queue. For the terminology within the paper we use:

\[ s = \text{service time} = \text{time spent by vehicles in the first position} = \frac{1}{c}, \text{if we represent the priority system by a M/M/1-queue} \]

\[ d = \text{delay} = \text{time spent by vehicles in the queue (without the first position)} \]
w = waiting time = time spent by minor street vehicles in the priority system
= d + s (definition following Heidemann’s (2002) proposal)

We denote the averages by small letters, whereas capitals stand for the sum of all delays (D) and the sum of all waiting times (W) respectively.

Figure 1: The „priority system“

There are different basic methods for estimating the capacity c, i.e. the maximum possible throughput for the minor stream, like the empirical regression theory (Kimber, Coombe, 1980), the critical gap theory, or - more recently - the conflict method (Brilon, Wu, 2002; Brilon, Miltner, 2003). These methods become even more complex if lower rank movements (e.g. left turning minor street movement) are treated. A good overview is given by Troutbeck, Brilon (2000) or Luttinen (2004). These methods should not be discussed here in detail. It is, however, desirable that the estimation of delay should become independent from the way of capacity calculation. Otherwise the solutions will become too complicated.

For undersaturated conditions, i.e. demand q is less than capacity c, the approximation of the priority system by an M/M/1-queue is rather popular among researchers (cf: Kimber, Hollis, 1979, Kimber et al (1986), or Heidemann, 2002). Figure 2 shows that the M/M/1-delay is not necessarily equal to the priority system delay. For this example illustration the average waiting time has been calculated based on gap acceptance theory using a set of equations given by Kremser (1962, 1964) and arranged as an equation for average waiting time by Brilon (1988). This arrangement is based on Yeo's results (Yeo, 1962; see also comments in Brilon, 1995). Kremser’s equations have later been improved by Daganzo (1977). This improvement has not been used for this example due to the rather complicated form of the equations. For this comparison the following parameters have been chosen: t_c = critical gap = 6 s, t_f = follow-up time = 3 s, q_p = major street volume = 350 veh/h (Figure 2, left side) and q_p = 600 veh/h (right side).

For the M/M/1-queue the average time of customers in the system (i.e. the waiting time) is

\[
\frac{w}{R} = \frac{1}{c - q} = \frac{1}{c \cdot (1 - x)}
\]

The average delay for the M/M/1-queue is
\[
d = w - \frac{1}{c} = \frac{x}{c \cdot (1-x)}
\]

where

- \( w \) = average waiting time [s]
- \( d \) = average delay [s]
- \( R \) = reserve capacity = \( c - q \) [veh/s]
- \( c \) = capacity [veh/s]
- \( q \) = demand volume [veh/s]
- \( x \) = degree of saturation = \( q/c \) [-]

Figure 2  Comparison of the M/M/1-waiting time with the waiting time calculated by the Kremser/Brilon-method (K/B); left side: \( q_p = 350 \) veh/h, right side: \( q_p = 600 \) veh/h

Figure 2 shows that there are cases of complete compliance of average waiting time between the M/M/1-solution and the priority system. However, also significant differences might be possible where the difference could be even larger than Figure 2 suggests, depending on the driver's behavior characteristics (expressed by the critical gap \( t_c \) and the follow-up time \( t_f \)) and on major street traffic volumes. Nevertheless, due to the similarities there is no useful alternative than to represent the priority system by a M/M/1-queue, if parameters of traffic performance like delays or queue lengths are to be described. Thus, the use of rather complicated equations for steady state priority delay is avoided. Moreover, the delay estimation becomes independent from the method of capacity calculation. It should, however, be noted that the representation of the priority system by the M/M/1-queue can cause a bias, especially for large major traffic volumes, i.e. for small capacities.

THE COORDINATE TRANSFORMATION METHOD

With this approximation by the M/M/1-queue only undersaturated conditions (i.e. \( x = q/c < 1 \)) can be described. For practical purposes it is, however, also necessary to estimate delays for situations where the demand volume \( q \) is variable over time and where it could even exceed
the capacity $c$ during a specific peak period. To describe the average delay suffered by minor street drivers during such an oversaturated period no solutions which are based on an exact statistical theory are in use since they – if a solution would be available – would become too complicated. Instead, rather pragmatic approximations are in use.

These approximations go back to an idea by a researcher named Whiting who contributed much to traffic research but did not publish in his own name due to personal reasons (Kimber, Hollis, 1979, p. 6; Allsop, 1992). The idea was developed fully by Kimber, Hollis (1979). The method is also characterized as coordinate transformation by several authors. It is illustrated in Figure 3. There we see the M/M/1-waiting time (eq. 1) as a function of the degree of saturation ($= x = q/c$) and the deterministic delay $d_d$. The deterministic delay is valid for a D/D/1-queueing system. The idea for the approximate assumption is:

- For very low saturation of the system the time-dependent effects do not play a role since the relaxation time, during which the system adapts to a changing demand is very small compared to the duration $T$ of the peak period. Thus, the time-dependent solution will be very close to the stationary solution which is represented by the M/M/1-queue.

- For extreme oversaturation (i.e. large $x$ and long period $T$ of oversaturation) the randomness of the system becomes less important. Effects of randomness then constitute only a very small part of the total delay. Thus, the average delay approaches the deterministic delay.

Therefore, the solution for the average delay in the time-dependent system should be a transition between the steady-state delay (M/M/1) and the deterministic delay. This transition curve, since a solution determined by stochastic theory seems to be too complicated, is then based on an approximation.

Three kinds of approximation may be used. The solution for each approach is derived from one of the following cases.

A1. additive; x-axis (see Figure 3)  
\[ \alpha = \beta \]

A2. multiplicative; x-axis (see Figure 3)  
\[ \frac{\alpha}{I} = \frac{\beta}{x_d} \]

A3. additive; R-axis (see Figure 4)  
\[ \alpha' = \beta' \]
where
\[ x = \text{degree of saturation} = \frac{q}{c} \quad \text{[ - ]} \]
\[ R = \text{reserve capacity} = c - q \quad \text{[veh/h]} \]
\[ \alpha, \beta, \alpha^*, \beta^* : \text{parameters, see Figure 3 and 4} \]

A potential fourth case, a multiplicative R-based approach gives no sense.

Each of these assumptions, constituting the fundament for the approximation, reveals a specific solution. There is no reasoning to prefer one of these approximations with regard to the basic sophistication.

**DETERMINISTIC DELAY**

To use the method of approximation we need to describe the deterministic delay in Figure 3 and Figure 4. As the D/D/1 system we understand a queuing system where all customers arrive with a headway of \(1/q\) (\(q = \text{demand volume}\)) and where they are served with a constant service time of \(s = 1/c\) (\(c = \text{capacity}\)). For \(x = q/c < 1\) (i.e. \(R > 0\)) there are no delays for customers in such a system. The only time which they spend in such a system is the service time. The reason why the D/D/1-system is treated, is the fact that for a D/D/1-system also delays for a temporary oversaturation can be determined.

In each queuing system the sum of all delays is the area between the cumulative arrivals and cumulative departures each represented by their function over time. Figure 5 gives an illustration.

Figure 5a shows the demand volume \(q\) over time versus the capacity \(c\), which is assumed to remain constant here. We see that during a peak period of duration \(T\) the demand exceeds the capacity, whereas the demand is assumed to be zero before and after the peak period. Then the sum of all arrived and departed vehicles, each as a pattern over time, is given in Figure 5b with a maximum difference \(N_T\) at the end of the peak period. The delay \(d\) for a vehicle arriving at time \(t\) can be obtained as the horizontal difference between the two curves. That means: the area included between both curves is the sum of all delays. The vertical difference between both curves, i.e. the queue length, is given in Figure 5c. Since the area between the two curves in Figure 5b is equal to the area under the queue length curve in Figure 5c, the sum \(D\) of all delays is the area below the curve for the length of the queue (Figure 5c).

Simple geometric considerations within Figure 5 reveal:

\[ N_T = \text{Max} \left\{ \left( q - c \right) \cdot T = -R \cdot T = c \cdot (x - 1) \cdot T \right\} \]

\[ D = \frac{1}{2} \cdot c \cdot x \cdot (x - 1) \cdot T^2 \]

Then the delay per vehicle averaged over all arriving vehicles for the deterministic case is

\[ d = \frac{D}{q \cdot T} = \frac{(x - 1) \cdot T}{2} \]

(5)
With approximation A1 we get a time-dependent solution for delay as:

\[ \omega = \frac{T}{4} \cdot \left\{ (x - 1) + \sqrt{\left(\frac{8 \cdot z}{T \cdot c}\right)^2 + (x - 1)^2} \right\} \]

where

- \( \omega \): average delay \( d \) or average waiting time \( w \) \([\text{s}]\)
- \( D \): sum of total delay \([\text{s}]\)
- \( x \): degree of saturation \( = \frac{q}{c} \) \([-]\)
- \( T \): duration of the peak period \([\text{s}]\)
- \( q \): demand traffic volume \([\text{veh/s}]\)
- \( c \): capacity \([\text{veh/s}]\)
- \( z \): parameter \([-]\)

With \( z = 1 \) we get a function for \( \omega(x) \) which constitutes a transition from waiting time \( w \) for undersaturated conditions to delay \( d \) for large \( x \). On the other side, for \( z = x \) eq. 6 converts into the well-known delay equation by Akcelik-Troutbeck (1991). This can be derived as a transition from random delay \( d \) for small \( x \) to deterministic waiting time \( w \) for over-saturation. Since the approximation converges on a term of different nature for both sides of the function (different for \( x \rightarrow 0 \) and for \( x \rightarrow \infty \)) this kind of solution shows a significant degree of inconsistency. This problem could be solved with the set of assumptions:
In addition, it is worth to notice that eq. 6 for any parameter $z$ is only valid for the unrealistic case of no arriving traffic before and after the peak period as well as constant capacity.

The more general case for the traffic demand pattern over time is illustrated in Figure 6. For these more general circumstances also the capacity $c$ is varied by a stepwise function. In addition an initial queue of length $N_0$ is assumed. Another difference to the previous case is that also after the observed peak period of duration $T$ a continued traffic demand $q_1$ is assumed. Since we concentrate on peak intervals with a potential temporary oversaturation the conditions

$$0 \leq q_1 << c_1 \quad q_1 < q$$

should be valid.

Two different clearance times for the queue can be defined. Period $a$ is the time after which the last vehicle arriving during the peak departs. $a_1$ is the time after which the expected length of the deterministic queue becomes zero.

![Figure 6: Deterministic delay for the general case.](image)

Again, also in Figure 6 (like in Figure 5) the total delay is the area between the two cumulative curves. There are, however, several possibilities which part of the area should be regarded as the relevant sum of delays. Here, different cases for the deterministic delay $D_D$ can be distinguished (table 1). As a general formula for the sum $D_D$ of all delays we can use:

<table>
<thead>
<tr>
<th>target function</th>
<th>$x &lt; 1$</th>
<th>$x &gt; 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega = \text{delay } d$</td>
<td>$z = x$</td>
<td>$z = 1$</td>
</tr>
<tr>
<td>$\omega = \text{waiting time } w$</td>
<td>$z = 1$</td>
<td>$z = x$</td>
</tr>
</tbody>
</table>
\[ D_D = \begin{cases} \frac{1}{2} \cdot [N_0 \cdot T + N_T \cdot (T + a)] & \text{for } N_0 > T \cdot c \cdot (1 - x) \\ \frac{N_0^2}{2 \cdot c \cdot (1 - x)} & \text{elsewhere} \end{cases} \] (8)

\[ N_T = \max \left\{ \begin{array}{c} N_0 + c \cdot (x - 1) \cdot T \\ 0 \end{array} \right. \] (9)

where

\[ N_T = \text{maximum queue length (deterministic)} \quad \text{[ veh ]} \]
\[ N_0 = \text{initial queue length} \quad \text{[ veh ]} \]
\[ a = \text{time to dissolve the queue due to oversaturation} \]
\[ \text{(according to table 1; cf. Fig. 1b and Fig. 6b)} \quad \text{[ s ]} \]

Table 1: Cases for the classification of total delay

<table>
<thead>
<tr>
<th>case</th>
<th>area in Figure 6b</th>
<th>equation for the term a (for the calculation of $D_D$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>A B C F</td>
<td>$a = 0$</td>
</tr>
<tr>
<td>D2</td>
<td>A B C E F</td>
<td>$a = \frac{N_T}{c_i}$</td>
</tr>
<tr>
<td>D3</td>
<td>A B C D E F</td>
<td>$a = a_i = \frac{N_T}{(c_i - q_i)}$</td>
</tr>
</tbody>
</table>

Each case has advantages for specific purposes:

D1  This definition restricts the consideration on delays which do occur exactly during the relevant peak interval. This definition is the only one to be applied when delays from successive intervals are added, e.g. over all hours of a whole day.

D2  This is the delay, which engineers usually define, when they estimate delays by empirical methods. It avoids to integrate delays experienced by vehicles arriving after the peak. But the D2-definition contains delays experienced after the end of the peak period. For the special case of $N_0 = 0$ this definition is closely related to the solution of eq. 6.

D3  This is the total delay which is induced into the system by the temporary overload. But it contains delays experienced after the end of the peak. Even delays for vehicles arriving after the considered peak period are involved into the total delay. For an economic assessment of delays, caused by specific peak periods, this definition is the preferential one, since it represents the total consequences of the overload happening during the peak period.

It should be noted again that $D_D$ covers delays – not waiting times.
From the sum $D_D$ of total deterministic delay the average delay $d_D$ has to be derived relating $D_D$ to those vehicles which are exposed to become involved into the queue of waiting vehicles. This means

$$d_D = \frac{D_D}{N}$$

where

- $d_D$ = average deterministic delay [s]
- $D_D$ = sum of all deterministic delays [s]
- $N$ = number of vehicles exposed to contribute to $D_D$ [-]

For undersaturated conditions (i.e. $x < 1$) $N$ contains all vehicles which arrive during the relevant time period $T$ with the consequence that $N = q \cdot T$.

For temporary oversaturation the calculation of $N$ is not self-evident. Then $N$ could comprise all vehicles which could be affected by the queue which is formed due to the oversaturation. Therefore, three cases can be formulated for the derivation of $N$:

- **N1:** $N$ contains those vehicles arriving during the time interval $(t, t + T)$; i.e. $N = q \cdot T$
- **N2:** $N$ contains vehicles which arrive during the time of an existing queue; i.e. $N = q \cdot (T + a)$
- **N3:** idem; $N = q \cdot (T + a_i)$

Of course, case N2 can only be combined with D2 and case N3 gives only sense with D3, whereas N1 can be combined with case D1, D2 and D3. Here it is preferred to relate all delays to the vehicles arriving during the peak period of duration $T$; i.e. we restrict ourselves to case N1.

One example for the deterministic delay depending on the degree of saturation $x$ is shown in Figure 7. In Figure 8 the deterministic delay is shown how it depends on the reserve capacity $R$. In both figures the limiting case

$$X = x_g = 1 - \frac{N_0}{T \cdot c} \rightarrow R = R_g = \frac{N_0}{T} \quad \text{with} \quad d = \frac{T \cdot N_0}{2 \cdot (T \cdot c - N_0)}$$

is marked. This is the point which - as a maximum - enables a dissipation of the initial queue (length $N_0$) within the peak period of duration $T$. Beyond this point (i.e. for $x > x_g$ or for $R < R_g$) the queue at the end of the peak period will be $> 0$ for the deterministic system. We see that $N_0$ has an influence on the shape of the curves for small $x$ (i.e. $x < 1$). For $N_0 > 0$ the value of $d$ increases to infinity for $x \to 0$, since due to $N_0$ there is always some delay experienced in interval $(t, t+T)$. With eq. 10, for small $x$ the number of $N$ is also small. This will need some special treatment later in this paper. Only for $N_0 = 0$ the deterministic delay starts from the point $(x = 1, d_0 = 0)$. For case D2 the relation $D_D = Funktion(x)$ is always nearly linear. D3 is identical with D2 for $q_1 = 0$. With increasing $q_1$ the D3-curve becomes increasingly concave. In the limiting case of $q_1 = c_1$ the curve for case D3 grows to infinity at $x = 1$. The D1-curve is always convex.
Figure 7: Deterministic delay as a function of x. For this example N₀ has been chosen as 20 (left side) and 0 (right side; both figures: c = 600 veh/h; c₁ = 600 veh/h; q₁ = 550 veh/h; T = 1 h).

Figure 8: Deterministic delay as a function of the reserve capacity R. For this example N₀ has been chosen as 20 (left side) and 0 (right side; both figures: c = 600 veh/h; c₁ = 600 veh/h; q₁ = 550 veh/h; T = 1 h).

It would now be desirable to conduct each of the approximations A₁ - A₃ for every definition D₁ to D₃ of deterministic delay. This may, in principle, be a possible option. In practice this is, however, nearly impossible and it gives no real sense. The reason is: Equations 8 and 9 combined with table 1 reveals rather complicated equations. For the approximations these equations have to be solved for x (and R in case A₃). This will give extremely complicated solutions. Then, from the equations in table 1 we get another set of equations which have to be solved for the delay d. The resulting equations will become even more complex, if all variables are used. It is, however, not very reasonable to develop extremely complex solutions if the result is still only an approximation.

Therefore, as a first approach we concentrate on the simple case where N₀ = 0 (i.e. no initial queue) and c₁ = c (i.e. constant capacity). This simplified case is not relevant for case D₁ and D₃. Thus, first of all we can derive equations representing the average delay for case D₂, which then will be used as a reference.

case D₂ + A₁:

\[
d = \frac{1}{4 \cdot c} \left[ 2 + c \cdot T \cdot (x - 1) + \sqrt{(2 + c \cdot T \cdot (x - 1))^2 + 8 \cdot x \cdot c \cdot T} \right]
\]

(12)

This equation should be identical with the Akcelik-Troutbeck equation, which is not the case. The reason is: this equation is derived from the delay d, both within the M/M/1-queue and
within the deterministic system. The original Akcelik-Troutbeck equation is derived as a transition from waiting time \( w \) in the M/M/1-case to the delay \( d \) in the deterministic case (see text in connection with eq. 6).

case D2 + A2:

\[
d = \frac{T}{4} \left[ x - 1 + \sqrt{(x-1)^2 + \frac{8 \cdot x}{c \cdot T}} \right]
\] (13)

The surprising result is that this equation - resulting from the multiplicative approximation A2 - is exactly the Akcelik-Troutbeck equation.

case D2 + A3:

\[
d = -\frac{1}{4 \cdot c} \left[ R \cdot T + 2 - \sqrt{(R \cdot T - 2)^2 + 8 \cdot c \cdot T} \right]
\] (14)

This equation is different from the corresponding solution by the author (Brilon, 1995, eq. 2.20). The reason is again that the older solution has been obtained by a transition from the waiting time \( w \) in the M/M/1-solution to the delay \( d \) in the deterministic case, which was not a consistent solution.

Even if these formulas look quite different from each other, the numerical results are rather similar. This is pointed out in Figure 9 for one example. But also with these small differences it is of interest which approach is the more realistic one. This is tested in the next section of this paper.

Figure 9: Average delay \( d \) as a function of the degree of saturation \( x \) (left side) and in relation to the reserve capacity \( R \) (right side). The figures compare the three kinds of approximation A1 – A3. Also the Markov chain results are indicated. Parameters for this example: \( N_0 = 0 \), \( c = c_1 = 600 \text{ veh/h} \).

**Analytical Solutions**

To check the desired approximate solutions for their correctness, methods for the determination of exact solutions for the time-dependent M/M/1-queue are desirable. Mathematical literature provides several solutions for the state probabilities or for the distribution function of the number of customers in the system also for dynamic (i.e. time dependent) queuing systems. One approach has been given by Heidemann (2002) by using LaPlace transformations for queue length distributions and delay distributions. With this
approach Heidemann confirmed the Akcelik-Troubeck solution. He did not, however, envision the other potential options for approximate equations. This analytical result contains the problem that no explicit formulas are used. Instead the LaPlace transforms have to be solved numerically to get applicable results.

Other candidates for equations to describe time dependent state probabilities within time dependent queuing systems can be found in Takacs (1960), Morse (1958/1976), or Tarabia (2000). They offer rather complicated equations containing trigonometric functions (sin, cos, …). The critical point for all of these solutions is, however, that they do not allow input volumes q exceeding the capacity at any time. Thus, for temporary oversaturation the conditions for the analytical derivations are not fulfilled. As a consequence these analytical solutions are only of limited usefulness for application in traffic engineering, since here a temporary oversaturation of the system is the crucial case for application.

Newell (1982) proposed the so-called diffusion theory to estimate the average queue length as a pattern over time for a given capacity and demand pattern. Also the standard deviation of queue lengths can be determined. Troutbeck, Blogg (1998) have tested this approach. They confirm the quality of Newell’s formula based on comparisons to stochastic simulations. However, they also underline the approximate nature of Newell’s solution which produces biased results for small queue lengths. Troutbeck, Blogg do also identify the limits of the solution for the time dependent queue length estimation according to Kimber et al (1986). In any case all analytical solutions aim at an estimation of queue length from which average delays must be determined. Non of these approaches claims for an analytically exact solution neither for queue length nor for delays. It is, however, not desirable to check one approximate solution against another approximation as a reference.

**MARKOV-CHAINS**

Even if an analytical solution is not visible, there are possibilities to get exact results for the delay within each queuing system by using numerical methods. These are based on the Markov-properties of the priority process which are especially valid for the M/M/1-approximation.

The average number of vehicles in the system can be estimated by the following procedure. We observe the queuing system in intervals of \( \Delta t = 1 \) minute duration. Since arrivals and possible departures are Poisson, we get

\[
\begin{align*}
    a_i(t) &= e^{-q(t) \Delta t} \frac{(q(t) \cdot \Delta t)^i}{i!} \\
    b_i(t) &= e^{-c(t) \Delta t} \frac{(c(t) \cdot \Delta t)^i}{i!}
\end{align*}
\]

where

\[
\begin{align*}
    a_i(t) &= \text{probability of i arrivals during the interval (t, t+\Delta t)} \quad [-] \\
    b_i(t) &= \text{probability that i departures are possible during the interval (t, t+\Delta t)} \quad [-] \\
    q(t) &= \text{traffic volume} \quad [\text{veh/min}]
\end{align*}
\]
\(c(t) = \) capacity \([\text{veh/min}]\)

\(q(t)\) and \(c(t)\) are assumed to remain constant with sufficient degree of approximation during the interval \((t, t + \Delta t)\).

\(\Delta t = \) duration of the time interval (here: 1 minute). We assume that \(T = k*\Delta t\) where \(k\) is any integer number.

From the \(a_i\) and \(b_i\) we form two quadratic matrices \(\overline{A}\) and \(\overline{B}\) with the dimension \(n\).

\[
\overline{A}(t) = \begin{bmatrix}
a_{0}(t) & a_{1}(t) & a_{2}(t) & a_{3}(t) & \ldots \\
0 & a_{0}(t) & a_{1}(t) & a_{2}(t) & \ldots \\
0 & 0 & a_{0}(t) & a_{1}(t) & \ldots \\
0 & 0 & 0 & a_{0}(t) & \ldots \\
\vdots & \vdots & \vdots & \vdots & \ddots \\
\end{bmatrix}
\]

\(\overline{B}(t) = \begin{bmatrix}
1 & 0 & 0 & 0 & \ldots \\
1-b_{0}(t) & b_{1}(t) & 0 & 0 & \ldots \\
1-b_{0}(t)-b_{1}(t) & b_{2}(t) & b_{1}(t) & 0 & \ldots \\
1-b_{0}(t)-b_{1}(t)-b_{2}(t) & b_{3}(t) & b_{2}(t) & b_{1}(t) & b_{0}(t) & \ldots \\
\vdots & \vdots & \vdots & \vdots & \ddots \\
\end{bmatrix}
\)

Then the transition probabilities are given by the matrix \(\overline{P} = \overline{A} \cdot \overline{B}\). This means: each term \(p_{ij}\) of the matrix \(\overline{P}\) is calculated as

\[p_{ij}(t) = \sum_{k=0}^{n} a_{ik}(t) \cdot b_{kj}(t)\]  

(18)

where

\(p_{ij}(t) = \) probability that the number of vehicle in the queue changes from \(i\) to \(j\) within the interval \((t, t + \Delta t)\)

\(a_{ik}(t) = \) term of the matrix \(\overline{A}\) at time \(t\) in row \(i\) and column \(k\)

\(b_{kj}(t) = \) term of the matrix \(\overline{B}\) at time \(t\) in row \(k\) and column \(j\)

\(n = \) number for the range of numerical calculations. \(n\) should be significantly larger than the maximum possible queue length.

Then the state probabilities of the system will change over time according to

\[p_{i}(t+1) = \sum_{k=0}^{n} p_{i}(t) \cdot p_{ik}(t)\]  

(19)

The equation makes it possible to calculate each of the state probabilities \(p_{i}\) at any time \(t + 1\) out of the previous state probabilities at time \(t\).

Then the average number \(N(t)\) of vehicles in the system at time \(t\) is

\[N(t) = \sum_{i=0}^{\infty} i \cdot p_{i}(t)\]  

(20)
This is a numerically exact calculation for the dynamics of the expected number of vehicles in the system. Again – like in the case of the deterministic system – the area below the curve for \( N(t) \) is the sum of all delays.

\[
D = \int_0^\infty N(t)dt
\]  

(21)

Or with stepwise constant demand \( q(t) \) and capacity \( c(t) \):

\[
D = \sum_{t=0}^{t_{\text{max}}} N(t) \cdot \Delta t
\]  

(22)

This sum of delays then can be related to the number of arrivals during the peak period of duration \( T \) to estimate the average delay \( d \).

\[
d = \frac{\sum_{t=0}^{t_{\text{max}}} N(t) \cdot \Delta t}{\sum_{t=0}^{t_{\text{max}}} q(t) \cdot \Delta t} = \frac{\sum_{t=0}^{t_{\text{max}}} N(t)}{\sum_{t=0}^{t_{\text{max}}} q(t)}
\]  

(23)

Here \( t_{\text{max}} \) is the number of the final interval (duration \( \Delta t \)) where \( N(t) \) becomes 0 if case D3 (Table 1) is applied. For case D1 \( t_{\text{max}} = T / \Delta t \). Case D2 can be constructed by setting \( q(t) = q_1(t) = 0 \) after time \( T \).

### Table 2: Comparison of results for average delay \( d \): Residual standard deviation for Case D2, Cases A1 - A3 and eq. 6

<table>
<thead>
<tr>
<th>compared to</th>
<th>capacity</th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>eq. 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Markow-chain calculation 1)</td>
<td>c = 100</td>
<td>14.63</td>
<td>13.24</td>
<td>24.38</td>
<td>15.75</td>
</tr>
<tr>
<td></td>
<td>c = 300</td>
<td>9.36</td>
<td>4.31</td>
<td>3.07</td>
<td>2.02</td>
</tr>
<tr>
<td></td>
<td>c = 600</td>
<td>7.84</td>
<td>5.38</td>
<td>3.58</td>
<td>4.02</td>
</tr>
<tr>
<td>compared to simulation 2)</td>
<td>c = 100</td>
<td>81.7</td>
<td>76.8</td>
<td>70.8</td>
<td>78.0</td>
</tr>
<tr>
<td></td>
<td>c = 300</td>
<td>42.9</td>
<td>40.8</td>
<td>39.3</td>
<td>41.1</td>
</tr>
<tr>
<td></td>
<td>c = 600</td>
<td>28.0</td>
<td>27.4</td>
<td>27.1</td>
<td>27.8</td>
</tr>
</tbody>
</table>

1) values = standard deviations for the difference between the result from eq. 12 - 14 and the Markow-chain result; compared over each \( x \in \{0.5 \text{ (0.05) 1.2}\} \)

2) values = standard deviations for the difference between results from eq. 12 - 14 and the simulated average delay; compared for 1000 1-hour simulation runs with various combinations of \( q \) and \( c \) within the interval \( x = 0.5, 1.2 \)

Markow-chain calculations have been performed for several examples. Results for average delay according to case D2 were evaluated for several combinations of \( c \) and \( T \) with emphasis on longer peak periods in the range of \( T \sim 1 \text{ h} \). It turned out that on the scale of Figure 9 the relation for \( d = F(x) \) or \( d = F(R) \) estimated by eq. 12 – 14 matched quite well with Markow-chain results. The differences are so small that on the scale of Figure 9 they are not visible. To notice the differences between eq. 12 – 14 and Markow-chain results Figure 10 has been plotted. Here we see that the differences are quite small in absolute terms (upper part
of the figure). The differences may become quite significant if we treat them in relative terms (bottom part of the figure). But the extremely large relative differences for small x-values are in the area of nearly zero delays, so that here the absolute differences are very small. Even if a remarkable deviation of the estimated average delays from the exact values could occur, the degree of approximation seems to be quite acceptable for practical application. Similar figures can be plotted for other combinations of parameters. The tendency of the 3 curves remains similar, the absolute value of the differences may, however, vary. They become rather significant if the capacity has very low values.

In any case the A1- approximation (i.e. the additive approach over x) turns out to be of lowest quality. It is not very clear, which of the A2 or A3 is better. They seem to be equivalent with a small advantage for A3, the additive reserve capacity based solution. Also the solution for delays $d$ according to eq. 6 shows a rather good performance.

**Figure 10:** Difference in average delay $d$ between Markov-chain results and the approximate solutions A1 - A3 as a function of the degree of saturation x. Parameters for this example: $T = 1 \text{ h}; N_0 = 0, c = c_1 = 600 \text{ veh/h}$ (left side) and $c = c_1 = 300 \text{ veh/h}$ (right side).

### SIMULATION

Another method for comparing the approximate results with "true" values is to use stochastic simulation. A computer program for simulating the priority system with constant critical gaps $t_c = 6 \text{ s}$ and $t_f = 3 \text{ s}$ has been written where drivers behaved in a consistent and homogeneous
way. The program evaluates the average delay in peak periods of duration \( T \) (here \( T = 1 \) h) according to approach D2. A large number of repetitions is possible. Figure 11 shows that the delays do - on average - follow the calculated curves. The results for the 1-hour average of delay do, however, vary over a quite remarkably wide area. The standard deviation of delays (standard deviation between average delays over 1-hour intervals) is always in the range of 0.7 of the mean of the average delays.

The three curves represent the approximations for case A1 (upper curve, eq. 12), A2 (eq. 13), and A3 (lower curve, eq. 14). The residual standard deviations for the simulated points (relative to eq. 12 – 14) are also given in Table 2. If we try to interpret the small differences the results support the A3-solution. The difference to the A2-solution is, however, only quite small. With respect to the large variance of the average delay such small differences become meaningless in practical terms. Also eq. 6 provides an adequate degree of approximation.

All equations compared to simulation results show a tendency to slightly overestimate average delays in the range of \( x = 1 \). The validity of all equations might be improved if for the basic equations of approximation (cases A1 to A3) a factor would be used relating the values of \( a \) and \( b \) (or \( a^* \) and \( b^* \)) against each other. This may be subject of further research.

At this point it can be stated, also on the background of more example calculations, that the approximation method A3 gives the best correlation to simulation and Markov-chain results. The method A2 is, however, very close up. Based on this experience and knowing that the same result must not necessarily be obtained also for cases D1 and D3, method A1 is not further treated here.

---

**Figure 11:** Simulation results for the priority system. Each point represents the average delay over a 1-hour-peak period according to case D2 in relation to the degree of saturation \( x \). The larger points represent the standard deviation between the results for average delay in the corresponding range of \( x \). Parameters for this example:

\[ T = 1 \text{ h}; N_0 = 0, c = c_1 = 600 \text{ veh/h (left side)} \] and \( c = c_1 = 300 \text{ veh/h (right side)} \).

**EFFECT OF \( N_0 \)**

Up to now we have studied the unrealistic simplified case of \( N_0 = 0, c_1 = c, \) and \( x_1 = q_1/c_1 = 0 \). To adjust the solution to more realistic circumstances \( N_0 \), the initial queue length, should be allowed to have any positive integer value. Then the sum of deterministic delays assumes a
function over $x$ or over $R$ like it is illustrated in Figure 12. The direct application of the principle of approximations A1 - A3 does not lead to useful results. Besides the fact that the equations assume unreasonable complicated functions, there is also the problem that the average delay does not only increase for large $x$ (or small $R$) but also in the vicinity of $x \to 0$ (i.e. $R \to c$) due to eq. 10. If $x = 0$ the term $N = q \cdot T = x \cdot c \cdot T$ in the numerator of eq. 10 becomes 0, such that with

$$D_0 = \frac{N_0^2}{2 \cdot c} \quad \text{(24)}$$

where

$$D_0 = \text{sum of delays for the case of no traffic arriving during the interval } (t, t+T)$$

there will be a minimum sum of delay in the denominator. Such a function is not accessible to approximation A1 - A3.

One solution might be to perform the same type of approximation on the scale of $D$, i.e. with the sum $D$ of delays instead of the average individual delay $d$. Trying this, the result is quite discouraging because the equations become unacceptably complicated including irrational functions. Nevertheless, a useful solution was found via the treatment of the sum of delays calculated from eq. 13 and 14 by multiplication with $N = q \cdot T = x \cdot c \cdot T$, which is the number of vehicles arriving during the peak interval $(t, t+T)$. In addition to this sum of delays, the minimum amount of delays $D_0$ which goes back to the initial queue, has to be added. It can be obtained from Figure 7, that also the asymptote for the deterministic delay has to be transformed from $d_0 = T/2(x-1)$ (in case $D_2; N_0 = 0$) to $d_0 = T/2(x-\hat{x})$ where

$$\hat{x} = l - \left( \frac{l}{T \cdot c} + \frac{l}{T \cdot c - N_0} \right) \cdot N_0 \quad \text{(25)}$$

Thus, the sum of delays can be described with a rather good approximation by the equation

$$D = D_0 + \frac{T^2}{4} \cdot c \cdot x \cdot \left\{ \Delta x + \sqrt{\Delta x^2 + \frac{8}{c \cdot T} \cdot x} \right\} \quad \text{(26)}$$

where

$$\Delta x = x - \hat{x}$$

The similar result can be obtained by approximation A3 using the reserve capacity $R$.

$$D = D_0 - \frac{(c-R) \cdot T}{4 \cdot c} \cdot \left( \Delta R \cdot T + 2 - \sqrt{(\Delta R \cdot T - 2)^2 + 8 \cdot c \cdot T} \right) \quad \text{(27)}$$

where

$$\Delta R = R - \hat{R}$$

$$\hat{R} = \frac{N_0 \cdot (2 \cdot T \cdot c - N_0)}{T \cdot (T \cdot c - N_0)}$$

Then the average delay $d$ is calculated from eq.10 with $D$ from eq.26 or 27 and $N = q \cdot T = x \cdot c \cdot T$. Figure 12 shows the resulting average delay $d$ for one example. This picture and other examples show: There is a rather good correspondence between the approximate formulas and the exact values represented by Markov-chain results. Only in the
range of $x = 0.3$ to 0.7 there are smaller differences which increase if $N_0$ grows into unrealistic large values (e.g. above 30 vehicles). In any case the preciseness is quite sufficient for practice. For larger $x$ the degree of approximation is quite good for any $N_0$-value.

![Graphs showing average delay for different values of $x$ and $R$.](image)

Figure 12: Average delay according to eq. 26 for an initial queue of $N_0 = 20$ depending on $x$ and according to eq. 27 depending on $R$. Parameters for this example: $T = 1$ h; $N_0 = 20$, $c = c_1 = 600$ veh/h. The figures on the left show results for the full scale of $x$ and $R$ whereas the right side is focusing on the more interesting area close to $x = 1$ ($R = 0$).

**EFFECTS OF THE POST-PEAK PERIOD**

There is still one case which has not yet been solved. This is the influence of the capacity and demand, which is prevailing after the peak period and which has an influence on the delay for the vehicles arriving during the peak interval $(t, t + T)$ in cases D2 and D3.

At first, for case D2 the effect of the post-peak capacity $c_1$ (which may be different from the peak period capacity $c$) can be taken into account by adding the difference in deterministic delay to the sum of all delays. As a consequence, the sum of delays (eq. 26 and 27) has to be modified as

$$D_{c_1} = D - \frac{N_T^2}{2} \left( \frac{1}{c} - \frac{1}{c_1} \right)$$

(28)
where

\[ D_{c1} = \text{sum of delays for } c_1 \neq c \quad [s] \]
\[ D = \text{sum of delays for } c_1 = c \quad (\text{eq. 26 or 27}) \quad [s] \]
\[ q_1 = \text{post peak traffic demand} \quad [\text{veh/s}] \]
\[ c_1 = \text{post peak capacity} \quad [\text{veh/s}] \]
\[ N_T = \text{max. deterministic queue length at the end of the peak period (cf. eq. 10)} \quad [-] \]

In the similar way the additional sum of delay which is due to definition D3 is given by

\[ D_{D3} = D_{c1} + \frac{N_T^2}{2 \cdot c_1} \cdot \frac{q_1}{c_1 - q_1} = D_{c1} + \frac{N_T^2}{2} \cdot \frac{x_1}{c_1 \cdot (1 - x_1)} = D_{c1} + \frac{N_T^2}{2} \cdot \frac{c_1 - R_1}{R_1 \cdot c_1} \quad (29) \]

where

\[ x_1 = \text{degree of saturation after the peak period} = \frac{q_1}{c_1} \quad [-] \]
\[ R_1 = \text{reserve capacity after the peak period} = c_1 - q_1 \quad [\text{veh/s}] \]

Table 3: Formulas for application

<table>
<thead>
<tr>
<th></th>
<th>A2 (using ( x = \frac{q_1}{c_1} ))</th>
<th>A3 (using ( R = q_1 - c_1 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>( D_1 = D_0 + \frac{c \cdot T^2}{4} \left( x - 1 \cdot \frac{2 \cdot x}{c \cdot T} + \sqrt{(x - 1)^2 + \frac{4 \cdot x}{c \cdot T} \left(1 + x + \frac{x}{c \cdot T}\right)}\right) )</td>
<td>( D_1 = D_0 + \frac{(c - R) \cdot T}{4 \cdot c} \left( \Delta R \cdot T + 2 - \sqrt{(\Delta R \cdot T - 2)^2 + 8 \cdot c \cdot T}\right) )</td>
</tr>
<tr>
<td></td>
<td>( \Delta x = x - 1 + \left( \frac{1}{c \cdot T} + \frac{1}{c \cdot T - N_0} \right) \cdot N_0 )</td>
<td>( \Delta R = R - N_0 \cdot \frac{2 \cdot c \cdot T - N_0}{T \cdot (c \cdot T - N_0)} )</td>
</tr>
<tr>
<td></td>
<td>( a = \frac{N_T}{c_1} )</td>
<td></td>
</tr>
<tr>
<td>D2</td>
<td>( D_{D2} = D_{26} + \frac{T^2}{4} \cdot c \cdot x \left( \Delta x + \sqrt{\Delta x^2 + \frac{8 \cdot x}{c \cdot T}} \right) )</td>
<td>( D_{27} = D_0 - \frac{(c - R) \cdot T}{4 \cdot c} \left( \Delta R \cdot T + 2 - \sqrt{(\Delta R \cdot T - 2)^2 + 8 \cdot c \cdot T}\right) )</td>
</tr>
<tr>
<td></td>
<td>( \Delta x = x - 1 + \left( \frac{1}{c \cdot T} + \frac{1}{c \cdot T - N_0} \right) \cdot N_0 )</td>
<td>( \Delta R = R - N_0 \cdot \frac{2 \cdot c \cdot T - N_0}{T \cdot (c \cdot T - N_0)} )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \Delta R = R - N_0 \cdot \frac{2 \cdot c \cdot T - N_0}{T \cdot (c \cdot T - N_0)} )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( a = \frac{N_T}{c_1} )</td>
</tr>
<tr>
<td>D3</td>
<td>( D_3 = D_{26} + \frac{N_T^2}{2 \cdot c_1} \cdot \frac{x_1}{c_1 \cdot (1 - x_1)} )</td>
<td>( D_3 = D_{27} + \frac{N_T^2}{2} \cdot \frac{c_1 - R_1}{R_1 \cdot c_1} )</td>
</tr>
<tr>
<td></td>
<td>( a = \frac{N_T}{c_1 \cdot (1 - x_1)} )</td>
<td>( a = \frac{N_T}{R_1} )</td>
</tr>
<tr>
<td></td>
<td>( N_T = N_0 + c \cdot (x - 1) \cdot T )</td>
<td>( a = \frac{N_T}{R_1} )</td>
</tr>
<tr>
<td>d</td>
<td>( d = \frac{D}{x \cdot c \cdot T} )</td>
<td>( d = \frac{D}{(c - R) \cdot T} )</td>
</tr>
<tr>
<td>w</td>
<td>( w = d + \frac{T + a}{c \cdot T + c_1 \cdot a} )</td>
<td></td>
</tr>
</tbody>
</table>
Finally, the whole set of equations, which can be equally recommended based on approximations A2 and A3, is given in where

\[ x_1 = \text{degree of saturation after the peak period} = \frac{q_1}{c_1} \quad [-] \]

\[ R_1 = \text{reserve capacity after the peak period} = c_1 - q_1 \quad [\text{veh/s}] \]

Table 3 as an overview. The formulas in both columns of the table are alternative to each other. Here, the average waiting time \( w \) is calculated from the average delay \( d \) by adding the weighted average of the service times \( s = \frac{1}{c} = \text{peak period service time, weighted by } c \cdot T \), and \( s_1 = \frac{1}{c_1} = \text{post-peak service time, weighted by } c_1 \cdot a \).

**CASE D1**

To describe case D1 by the similar degree of precision it turned out to be insufficient just to correct the D2-results by some specific terms. Thus, just the similar derivations like for case D2 had to be performed with special attention to the D1-conditions. As result the following equations for the average delay can be given for case D1 – A2.

\[
D_1 = D_0 + \frac{c \cdot T^2}{4} \left( x - 1 - \frac{2 \cdot x}{c \cdot T} + \sqrt{(x-1)^2 + \frac{4 \cdot x}{c \cdot T} \left( 1 + x + \frac{x}{c \cdot T} \right)} \right) \quad (30)
\]

The solution for the case D1 – A3 is not possible. For the generalized case it leads to an undefined area in the vicinity of \( R = R^* \).

**EMPIRICAL EVIDENCE**

The measurement of average delays at unsignalized intersections is not a trivial task. For comparisons, due to the wide variance of delays, a large sample size is required. Thus, to get useful measurement data, long periods with constant traffic volumes would be needed which does hardly occur in reality. Moreover, it is not easy to find oversaturated priority intersections since under high traffic demand junctions usually are signalized. The observation of delays requires also a rather good overview since the end of the queue always has to be under control. Such observations have been performed and described by Brilon, Weinert (2002). In their sample there were 4 intersections (all T-junctions) at rural two-lane highways with a temporary overload. Here the left turner from the minor road (LTMR) was observed regarding delays. All the other movements were counted simultaneously. The comparisons here are made on the basis of 5-minute intervals (\( T = 5 \text{ minutes} \)). To estimate \( x \) in each time interval the capacity \( c \) for the LTMR was estimated on the basis of the method in the German HCM (HBS, 2001). The calculation of average delays is according to definition D2 – A2.

We see an agreement of empirical and calculated delays as it could be expected on the background of the wide variance of waiting times during relatively short time slices. Apparently, at point 28 (left upper part of Figure 13) the calculation fails, since for higher degree of saturation the measured delays remain much below the calculated values. The videos from the measurements, however, made clear that a remarkable amount of gap forcing takes place at this junction during overloaded periods which, of course, reduces delays significantly below the modeled results. On all the other points a sufficient correspondence between measured and calculated delays is observed. A closer coincidence can not be
expected, because of the large variation of delays and due to the fact that the estimated capacities must not necessarily represent the true maximum potential throughput during the observations correctly.

Figure 13: Comparison of measured waiting times with calculated values. The dark dots represent the average waiting time measured according to definition D2. The small circles represent the calculated waiting time for case D2-A2 for the same traffic data as during the measurement interval (T = 300 s).

OTHER CONSIDERATIONS

Of course, methods for the derivation of queue lengths are closely related to the subject of this paper. By Little’s well-known formula ($\lambda = q \cdot d$) there is a relation between the average delay $d$ and the average queue length $\lambda$. This formula is, however, only valid for stationary queues, i.e. for constant capacities $c$ and demand flows $q$ plus $x < 1$. As shown above (cf. Figure 5 and Figure 6) the queue length is not a steady state variable in case of time dependent demand and/or capacity and especially not during oversaturation, since then it is continuously growing. Thus, it gives only sense to describe the function $\lambda(t)$. To estimate this function several theoretical approaches have been published, e.g. Newell’s diffusion equations (Newell, 1982) or Kimber’s method (Kimber et al, 1986) (for comparisons see Troutbeck,
Blogg, 1998) where Newell’ method allows also to estimate standard deviations of the queue lengths based on approximate assumptions. For numerical calculations it is, however, more convenient to evaluate the exact results based on Markov-chain calculations. For one example this function $\lambda(t)$ is illustrated in Figure 14. In addition to the expected queue length (in veh; without the vehicle in service) also percentiles of the queue length are given as a function of time.

It can be seen that the linear shape of the queue length over time, like it is assumed in Figure 5 and Figure 6, is not realistic. Instead the function $\lambda(t)$ has a curved shape (cf. Troutbeck, Blogg, 1998). It is also questionable which parameter of these curves should be indicated by formulas for application in practice. Should it be the maximum of the percentile curves or some average of these curves? Estimations of queue length percentiles in practice are mainly based on Wu (1994) (cf. HCM, 2000 or HBS, 2001). It is imaginable that also queue length estimations might be accessible to a new consideration of systematic classification for time-dependent conditions.

Moreover, it will be desirable to extend the considerations in this paper to a forth case D4, in which the pattern of traffic demand over time has also a time-dependency during the peak period.

![Figure 14: Profile of the average queue length and the 50-, the 95-, plus the 99-percentile queue length for T = 1 h, c = c1 = 600 veh/h, q = 600 veh/h, q1 = 0, N0 = 0.](image)

**CONCLUSION**

Average delay at unsignalized intersections for temporarily oversaturated conditions is usually calculated by approximate equations. On the one hand these approximations, in principle, do very well match with more precise estimates for average delay. The set of possible solutions is, however, wider than usually assumed. The most commonly applied formulas are just one special case out of a set of equally valid solutions. It must also be noticed that the most popular solutions, like the Akcelik-Troutbeck equation, are only valid for rather significantly simplified conditions which are not too representative for real world conditions.
The systematic classification of definitions for the average delay (case D1 - D3) and of the sophistication for the approximation (case A1 - A3) leads to some differentiation among the complete set of possible solutions. As a consequence, each application of such a delay formula needs some understanding of the background in definitions.

Each numerical calculation of performance measures at unsignalized intersections is based on a representation of the "priority system" by an M/M/1-queue. This first step of approximation does not produce a perfect fit but it is justified as a solution sufficient for practical application, since any alternative would provide a lack of practicability.

For the time-dependent M/M/1-queue the average delay can be calculated exactly by a numerical evaluation of the Markov-chain concept. Comparisons with the approximate equations show that the conventional additive concept A1 for the approximation sophistication is not the best. The best fit is achieved for concept A3; i.e. an additive concept using reserve capacity $R$. This is closely followed by concept A2 (multiplicative concept using $x = \text{degree of saturation}$).

These results are also confirmed by stochastic simulation. Here, we also get information about the high values of standard deviations of delays with a coefficient of variation for the 1-hour average delay in a range of 0.7. Even if the precise measurement of average delays needs quite an effort, it can be shown that empiry is in coincidence with the theoretical results.

The extension of the simplified case D2 (no traffic before and after the peak period) is able to take into account the traffic conditions before (by the initial queue $N_0$) and after the peak period. Here formulas can be found which are only slightly more complicated than the traditional equations for the simple case. These approximate equations reveal a precision which is comparable to the traditional Akcelik-Troutbeck equation. Thus, they offer an extension of the traditional formulas into areas of more realistic conditions for practical application. Equations which are recommended for practical application are arranged in table 3. The Akcelik-Troutbeck formula receives an improvement by the use of eq. 6. This version is proposed instead of the addition of the service time (like in Akcelik, Troutbeck, 1991, eq. 5.4 - 5.6 and the HCM 2000, eq. 17-38).

As a result of the investigations a set of delay equations is available, which can be applied to estimate average delays at unsignalized intersections for well-defined traffic conditions. The paper makes clear that instead of an uncritical use of delay formulas a well sophisticated selection of the adequate equation is required also in practice.

REFERENCES


Kremser, H. (1964). Wartezeiten und Warteschlangen bei Einfädelung eines Poissonprozesses in einen anderen solchen Prozess (Delays and queues with one poisson process merging into another one). Oesterreichisches Ingenieur-Archiv (Austria), no. 18


