

## **Stochastic Incident Duration: Impact on Delay**

Submission date: November 10, 2009

Authors:

Victor L. Knoop MSc (corresponding author)

TRAIL Research School

Delft University of Technology

Stevinweg 1

2628 CN Delft

The Netherlands

Phone: +31 15 27 81723

v.l.knoop@tudelft.nl

Prof. dr. ir. Serge P. Hoogendoorn

TRAIL Research School

Delft University of Technology

Stevinweg 1

2628 CN Delft

The Netherlands

Phone: +31 15 27 81723

s.p.hoogendoorn@tudelft.nl

Prof. dr. Henk van Zuylen

TRAIL Research School

Delft University of Technology

Stevinweg 1

2628 CN Delft

The Netherlands

Phone: +31 15 27 81723

h.j.vanzuylen@tudelft.nl

Word count: 245 (abstract) + 5696 (body text) + (6 figures and tables)\*(250 words) = 7441 words

**Abstract**

The duration of incidents is a stochastic variable with a variation spread. This chapter analyzes the consequences of this stochastic nature of the duration in terms of delay. It uses shockwave theory to describe traffic states. As opposed to a point queue model, the head and the tail of the queue are separately modeled and in this way the spatial extent of the queue is properly described. Using the traffic states, the delay is analytically calculated.

The paper distinguishes between three scenarios: (1) an incident happens on a road stretch without any influence of a junction; (2) an incident happens upstream of a junction; a queue forms upstream of the incident and capacity of the downstream links is insufficient to handle the queue discharge rate; (3) an incident happens downstream of a junction and the tail of the queue crosses the junction.

We derive a formula for the total delay. Because the delay is a non-linear function of the duration, the expected delay is not equal to the delay of the incident with the expected duration. In the scenarios without spillback (the first two scenarios), the delay is proportional to the square of the duration of the blocking. The expected delay is expressed as a function of the variance of the duration of the blocking. Also, the variance of the average delay per involved traveler is expressed as function of the variance of the delay.

## 1 INTRODUCTION

According to Bates et al. (1) road users dislike unexpected delays even more than an expected delay. Models have been proposed to compute the delay caused by an incident. This paper continues on the models of Olmstead (2) and Li et al. (3). This paper will extend their models with the spatial extension of a queue or the stochasticity in incident duration or remaining capacity at the incident location.

It is important to compute the cost of the delay of an incident accurately. Besides other effects, the costs of delays play a role in deciding upon investment in for instance in safety measures and incident management. This paper will show that it is important to include network characteristics in computing delays. For each of the three general network layouts, an equation is formulated to compute the total delay caused by an incident with a variable duration. For this computation, only the average duration and the variability are needed; the shape of the distribution function is not needed. Also, the variability of the individual delay can be expressed as function of the average duration and the variance of the duration. This contribution shows (1) the importance of the duration of an incident (quadratic), which shift the road authority's focus towards shortening incident duration; (2) the importance of spillback queues; (3) a new method to compute the delays of incidents quickly, using shockwave theory.

This paper is set up as follows: first, a review is given of the literature on incident duration models and models that predict the delay caused by an incident. Section 3 describes the theory of queuing and explains the shockwave theory that we will use in this contribution. It also gives the formulas to compute the delay. The theory is applied to three test cases, that are a basic set of network elements, that are described in section 4. Section 5 shows the traffic states which are predicted by shockwave theory and computes the delays for deterministic durations and capacity reductions. Section 6 deals with the stochasticity and discusses how uncertainties are processed. Then, section 7 shows numerically the impact of the variance of the duration of the incident on the delay for real-life incident data. The concluding remarks can be found in section 8.

## 2 STATE OF THE ART

This section discusses the literature on two subjects. In section 2.1 the duration of incidents is discussed and how the variability of this duration can be reduced by classifying the incidents. Section 2.2 gives an overview of the literature on calculating the delay of incidents.

### 2.1 Incident duration prediction

To estimate the effects of an incident the moment at the moment it happens, it is worthwhile to have an indication of the duration. There are several models to estimate the incident duration.

A simple approach would be to take some (independent) observable variables and fit a linear-regression model on the duration of the incident (4, 5). For the Dutch motorways a linear regression model is fitted (6). These models show that fitting a model gives an indication of the incident duration, but the indications are not very accurate.

Wang et al. (7) predict the incident duration using two different models. Based on four input parameters (report mechanism, vehicle type involved, time of day, location), they predict the time a vehicle that breaks down remains at the same place using a fuzzy logic model and an artificial neural network. They conclude that the artificial neural network performs better with a root mean square error of 20 minutes and an  $R^2$  of 0.4. The models "had difficulties in predicting the outliers", by which they mean that they did not find a cause for the cases in which the duration was very long. They conclude that the errors might be due to imperfect or insufficient information about the incident type.

Boyles et al. (8) apply Bayesian techniques. They use many variables (44) that characterize an incident. The paper focusses on how characteristics of the incidents become available over time and how the incident prediction gets more accurate using more information. They use a naïve Bayesian model to classify the incidents in three duration groups: less than 30 minutes, 30-60 minutes and more than 60 minutes; the final decision to which group the incident belongs is made by a probability maximization for each of the groups. The paper shows a (small) increase of incidents which were classified in the right group, compared to a linear regression model. When information was added on the time at which information became available, the improvement increased.

Zhang et al. (9) propose to combine the linear regression model (6) with a classification tree. Compared to a normal linear regression model, they find a better prediction model. They mention explicitly that in 10% of the cases, the incident duration is much longer than the model predicts. Therefore, they neglect these cases. However, particularly these cases cause the largest delays. Therefore, in computing the duration we will keep the incident duration variable. This contribution will show that the variability of the incident duration causes a large part of the total delay.

## 2.2 Delays due to incidents

Fu and Rilett (10) discuss the influence of a stochastic incident duration on the delay. They conclude that using a mean value for the duration leads to an error in the delay. They show how one could use the probability density function to calculate the delay. They also show how this can be used on-line, during the incident, applying Bayesian theory. In their analysis they assume that the traffic jam does not occupy any space (vertical queuing model, Vickery (11)).

Olmstead (2) discusses the delay on the road as a result of an incident. He derives an equation for the delay for all travellers,  $D$ , in case of a deterministic duration  $\Delta T$  (and a vertical queue):

$$D(\Delta T) = \frac{1}{2} \frac{\Delta T^2 (C - rC) (Q - rC)}{C - Q} \quad (1)$$

The equation is presented using the symbols presented in table 1:  $C$  is the capacity,  $r$  the capacity reduction, and  $Q$  the demand. They all are kept constant throughout the incident duration,  $\Delta T$ .

Using this formula, he shows that the expected value of an incident with a stochastic duration is larger than the delay of an incident with the expected value of the duration. He proposes a additional term, linear with the variance of the duration,  $\text{Var}(\Delta T)$ , to compute the expected value (indicated with angle brackets) of the total delay of an incident with a *stochastic* duration:

$$\langle D \rangle = D(\langle \Delta T \rangle) + \frac{1}{2} \frac{(C - rC) (Q - rC)}{C - Q} \text{Var}(\Delta T) \quad (2)$$

The article mentions explicitly that this only holds for this vertical queuing model.

(3) show that the average delay per (delayed) traveler can be calculated using the average duration, but that variance of the delay per delayed traveler has the same problem as the total delay. They present the following equation to compute the variance of delay per delayed traveler  $A$ :

$$\text{Var}(A) = \frac{(Q - C)^2 \text{Var}(\Delta T)}{3Q^2} + \frac{(Q - C) \langle \Delta T \rangle}{12Q^2} \quad (3)$$

Note that this is the variance of the delay for the travelers that are delayed. The equation does not provide the number of travellers that encounter this delay.

Both articles assume vertical queues (or point queues). In this contribution the spatial dimension is also considered. We will analyse how these equations change if the queues have a spacial extent. The methodology used for this calculation is similar to the one proposed by Newell (12).

**TABLE 1** The symbols used

Symbol	Meaning
$L$	Length of a queue
$Q$	Demand ( <i>veh/h</i> )
$i$	link number
$\Psi_i$	Fraction of traffic turning to link $i$
$\psi$	Split fraction to link 2
$C_i$	Capacity on link $i$
$r$	reduction factor for capacity due to incident
$q$	Traffic flow
$k_2(q)$	Density of traffic, non-congested regime
$k_1(q)$	Density of traffic, congested regime
$\Delta T$	Duration of the incident
$\omega$	Shock wave speed
$x$	Position at the road
$t$	time
$t^{\text{end}}$	Time when the congestion at the downstream link ends (without intersection)
$x^{\text{end}}$	Position at which the congestion at the downstream link ends (without intersection)
$\Delta x$	Distance between the incident and the junction
$A$	Average delay per delayed traveler
$N$	Number of travelers delayed
$q_{\text{queue}}$	The rate at which drivers drive into a traffic jam

### 3 THEORY FOR MATHEMATICAL FORMULATION OF QUEUE LENGTHS

This section explains the queuing theory used in this contribution. The first subsections discusses the flow-density relationship that is used. The second subsection presents a mathematical expression for the speed of the boundary between phases. The third subsection presents a way to compute the delay from a space-time state diagram.

#### 3.1 Fundamental Diagram

The traffic flow modeling using shock wave theory assumes there are distinguished traffic states. A traffic state is uniquely characterized by the flow,  $q$  and the density  $k$ . A “fundamental diagram” poses the relationship between the density  $k$  and the flow  $q$ . We use a triangular fundamental diagram as proposed by Daganzo (13). This is a relatively simple fundamental diagram, but traffic can be modeled well using this shape.

$$\begin{aligned} q &= v_f * k & k < k_c \\ q &= -(k - k_j) \frac{C}{k_j - k_c} & k > k_c \end{aligned} \quad (4)$$

Two branches are separated; traffic states with a density lower than than the critical density  $k_c$  non-congested, traffic states with a density larger than  $k_c$  are congested.

For the sequel of this contribution, it is useful to rewrite equation 4 from  $q(k)$  to  $k(q)$ . For one flow, there are two different densities possible, a free flow density and congested density. We will refer to these

densities as  $k_2$  and  $k_1$  respectively. The relationship expressed in equation 4 now becomes

$$\begin{aligned} k_2 &= \frac{q}{v_f} = \frac{q}{\frac{k_c}{v_f}} && \text{free flow} \\ k_1 &= k_j + \frac{k_c - k_j}{C} q && \text{congested} \end{aligned} \quad (5)$$

### 3.2 Shock waves and their speeds

Suppose there are two different homogeneous traffic states, A and B, with properties  $\{q_A, k_A\}$  and  $\{q_B, k_B\}$  respectively. When describing the traffic situation with shock wave theory, these two states are separated by a boundary, referred to as ‘‘shock’’. The boundary between the two states propagates, by definition, with speed  $\omega_{AB}$ . The speed  $\omega_{AB}$  can be derived from the law of conservation of vehicles and the assumption that the flow depends on the density (see for instance (14)). It can be expressed as follows:

$$\omega = \frac{q_2 - q_1}{k_2 - k_1} \quad (6)$$

The rate at which drivers drive into a traffic jam,  $q_{\text{queue}}$  can be computed by multiplying the traffic density  $k$  with the speed at which they approach the traffic jam ( $v - \omega$ ):

$$q_{\text{queue}} = k(v - \omega) \quad (7)$$

This equation can also be used to compute the flow out of a traffic jam. The total number of travelers that drive into a traffic jam,  $N$ , now is the integral over time of equation (7):

$$N = \int q_{\text{queue}} dt = \int k(v - \omega) dt \quad (8)$$

### 3.3 Delays

A consequence of choosing a triangular fundamental diagram is that travelers only encounter delays when driving in a congested traffic state. At every state at the free branch of the fundamental diagram, the speed equals the free flow speed  $v_f$ .

We now compute the delay for a driver in traffic state A, who travels with speed  $v_A$  instead of the free speed  $v_f$ . We compute the delay he encounters is time  $dt$  assuming that in other time periods he travels without delay. The infinitesimal addition to the delay  $dd$  in an infinitesimal small time  $dt$  is:

$$dd = \frac{v_f - v_A}{v_f} dt \quad (9)$$

In total, there are  $N_A(t)$  drivers in traffic state A at moment  $t$ . The infinitesimal addition to the total extra delay at one moment in time caused by a traffic state now is the product of the number of drivers in that times the delay for a driver.

$$dD = N_A \frac{v_f - v_A}{v_f} dt \quad (10)$$

To find the total delay  $D$  for all drivers, the infinitesimal delays at  $dt$ , equation (10), needs to be integrated over time:

$$D = \int N_A(t) \frac{v_f - v_A}{v_f} dt \quad (11)$$

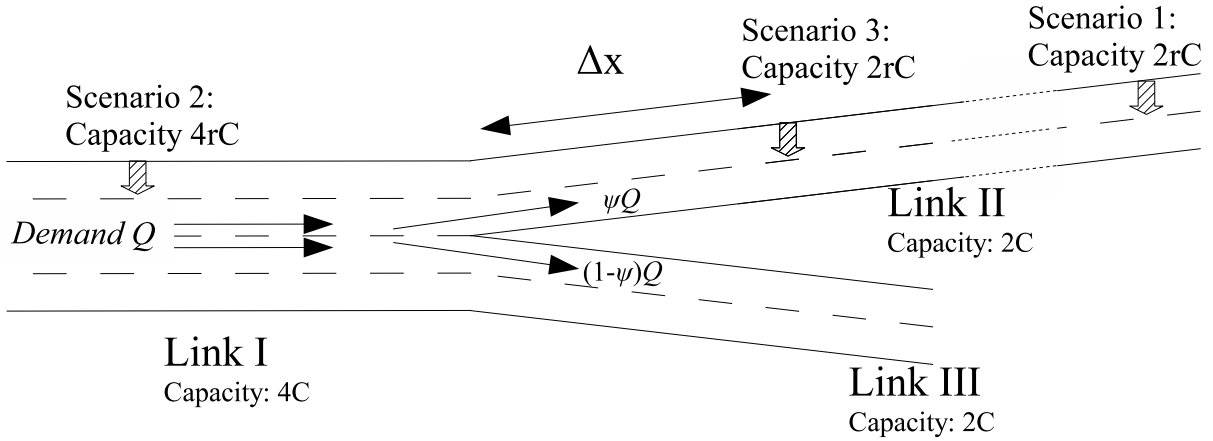


FIGURE 1 The used road layout - incident locations indicated by a shaded arrow

### 3.4 Relevance of modelling spatial extension

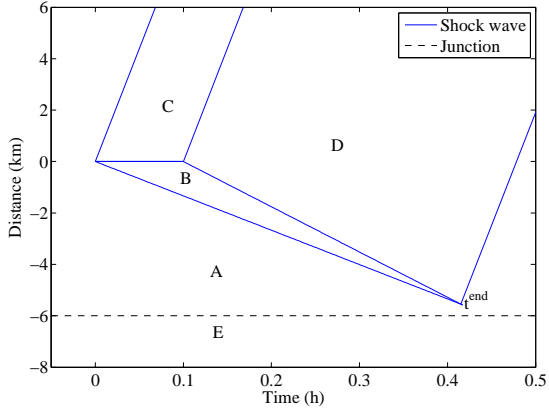
As long as the position of the tail of the queue does not influence the inflow pattern, the delay is determined by this inflow pattern and the outflow pattern. Therefore, in principle, one could calculate the delay using a vertical queuing model (11). This results in the correct results as long as the number of vehicles in the car is modelled correctly and the outflow is correct. For calculating the average delay it is also required that the number of vehicles and the number of unique vehicles in the queue are modelled correctly. This is not obvious with a vertical queuing model. Since for this property we need the full queuing dynamics, we will also compute the total delay using shockwave theory. Moreover, the paper then shows how the method can be used to compute the total delay.

## 4 ROAD LAYOUT FOR CONSIDERED CASES

The road layout considered shown in figure 1. Three different incident locations are considered, indicated with an shaded arrow. In case 1, discussed in section 5.1, there is a straight road with a capacity reduction. The next two cases use the network where traffic from 1 link have to split into 2 links. In case 2, discussed in section 5.2, a capacity is temporally reduced on the inflow link. After the incident has been cleared, The third case, discussed in section 5.3, has a temporal capacity reduction on one of the outflow links and the resulting queue grows further than the junction.

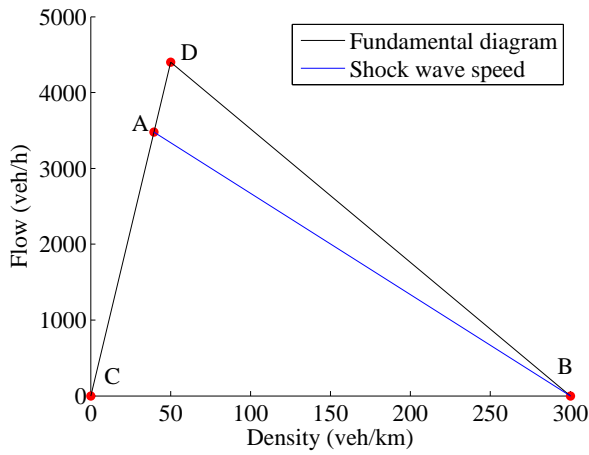
The road layout we chose for illustration purposes is a four lane freeway which splits into two times two lanes. For the sake of understanding of the concepts, the road properties are specified according to typical values. However, the mathematical formulae are derived for a general case where the capacities of the links are included as model parameters. Properties of the road are based on (15), the Dutch equivalent of the Highway Capacity Manual (16). For the numerical evaluations and graphs, we assume a capacity of 2200 *veh/h/lane*, a critical density of 25 *veh/km/lane*, a jam density of 150 *veh/km/lane*. We use a demand of 5800 *veh/h* of which 60% turns towards link 2.

The capacity at the location of the incident is considered as stochastic variable  $r$ .  $r$  is the fraction of the free capacity that is available during the incident, and varies from 0 to 1. Other network-configurations are basically a combination of the three road-layouts presented here.

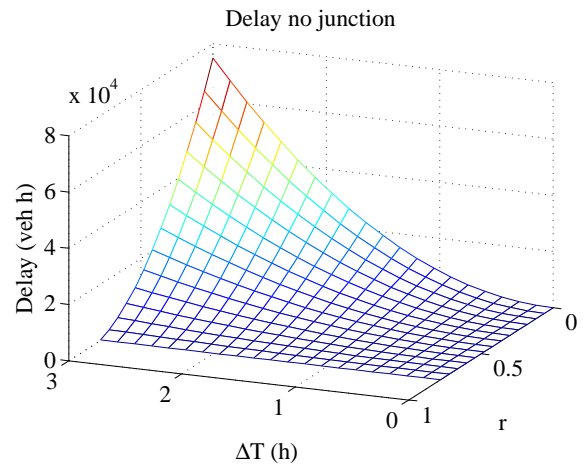


(a) Traffic states

State	Lanes	Flow	Congested
A	2	$\psi Q$	Free
B	2	$2rC$	Congested
C	2	$2rC$	Free
D	2	$2C$	Free
E	4	$Q$	Free



(b) The traffic states on the density-flow diagram

(c) Delay as function of  $r$  and  $\Delta T$ **FIGURE 2 Traffic situation for scenario 1**

## 5 APPLYING THE MODEL

This section describes the traffic states that are present for three different cases. Table 1 shows the symbols that are used throughout the section and at other places in the paper.

### 5.1 Scenario 1: no influence of junctions

#### 5.1.1 Traffic states

A typical pattern of traffic states is shown in figure 2a. For case in the figure, we choose  $r = 0$  and  $\Delta T = 0.1h$  and we assume the length of the link is long enough that the tail of the queue will not reach the end of the link. Therefore, a vertical queue would yield the same total delay as a dynamic model, but the queue dynamics and the average delay are different (see section 3.4). The same holds for a queue caused by a moving bottleneck. In the remainder of this section we will present the delay based on a calculation with complete queue dynamics for a fixed bottleneck, which is typically the case for an incident.

We choose the point of the accident to be at  $x = 0$  and at  $t = 0$ . During a time  $\Delta T$  the road is blocked at  $x = 0$ .



Lines mark the separation of different traffic flow regions. The demand at link 2 is  $\psi Q$ , which is the traffic flow in state A. The incident reduces the traffic flow to  $rC_2$ . Downstream of incident there is a free flow traffic state C with flow  $rC_2$ . Upstream of the incident, an area with congestion builds up (B). The flow there also is  $rC_2$ . After the incident has cleared, the outflow of the traffic jam is  $C_2$ , in a free flow state (D). The traffic states and shock waves are also indicated on the density-flow in figure 2b.

Using equation 6, we can compute the speed at which the boundary between traffic state A and B travels backwards,  $\omega_{AB}$ :

$$\omega_{AB} = \frac{Q\psi - rC_2}{k_2(Q\psi) - k_1(rC_2)} \quad (12)$$

Similarly, the boundary between traffic state B and D travels backwards with speed  $\omega_{BD}$ :

$$\omega_{BD} = \frac{Q\psi - C_2}{k_2(Q\psi) - k_1(C_2)} \quad (13)$$

This shock wave travels upstream with a larger absolute speed than the shock wave at the upstream end of the congested area. The travel position of the wave fronts in time and space are:

$$x_1(t) = \omega_{AB}t \quad (14)$$

$$x_2(t) = \omega_{BD}(t - \Delta T) \quad (15)$$

If the link is long enough, the congestion is solved when  $x_1$  equals  $x_2$  (at moment  $t^{\text{end}}$ )

$$\begin{aligned} \omega_{AB}t^{\text{end}} &= \omega_{BD}(t^{\text{end}} - \Delta T) \\ t^{\text{end}} &= \frac{\omega_{BD}\Delta T}{\omega_{BD} - \omega_{AB}} \end{aligned} \quad (16)$$

Substituting  $t^{\text{end}}$  in equation (14) then gives the position on the road where the two waves meet:

$$x^{\text{end}} = \frac{\omega_{AB}\omega_{BD}\Delta T}{\omega_{BD} - \omega_{AB}}; \quad (17)$$

During this scenario, we assume that congestion does not reach the junction upstream of the incident location. Having a value for  $x^{\text{end}}$ , this now can be quantified:

$$\begin{aligned} \Delta x &> x^{\text{end}} \\ \Delta x &> \frac{\omega_{AB}\omega_{BD}\Delta T}{\omega_{BD} - \omega_{AB}} \end{aligned} \quad (18)$$

### 5.1.2 Delay

To compute the delay, one first needs to know the number of vehicles in the traffic jam. This can be computed from the length of the queue in meters,  $L(t)$ , and the density in the congested area,  $k_1(rC_2)$ . In this case, the length of the queue  $L(t)$  is:

$$L(t) = x_2(t) - x_1(t) \quad (19)$$

The number of vehicles in this area,  $N_B$  is computed as:

$$\begin{aligned} lN_B &= L_B(t)k_1(rC_2) \\ N_B &= (x_2(t) - x_1(t))k_1(rC_2) \end{aligned} \quad (20)$$

For the simplicity of the computations, we split the time during which there is congestion ( $0 < t \leq t^{\text{end}}$ ) into two parts. We distinguish a first part in which congestion builds up ( $0 < t \leq \Delta T$ ), and a second part the congestion solves ( $\Delta T < t \leq t^{\text{end}}$ ). The respective queue lengths are:

$$\begin{aligned} N_B &= \omega_{AB}t & 0 < t \leq \Delta T \\ N_B &= \omega_{AB}\Delta T - (\omega_{AB} - \omega_{BD})(t - \Delta T) & \Delta T < t < t^{\text{end}} \end{aligned} \quad (21)$$

If we split equation (11) for the two time periods (queue growing and queue solving), we obtain the following:

$$D = \int_0^{\Delta T} N_B(t) \frac{v_f - v_B}{v_f} dt + \int_{\Delta T}^{t^{\text{end}}} N_B(t) \frac{v_f - v_B}{v_f} dt \quad (22)$$

Substituting the value for  $N_B$  (equation (21)) into this equation, gives the following

$$\begin{aligned} D &= \frac{1}{2} \frac{\omega_{AB}(v_f - v_B)\Delta T^2}{v_f} \\ &+ \frac{1}{2} \frac{(-\omega_{AB} + \omega_2)(v_f - v_1) \left( (t^{\text{end}})^2 - \Delta T^2 \right)}{v_f} \\ &+ \frac{(\omega_{AB}\Delta T + (\omega_{AB} - \omega_{BD})\Delta T)(v_f - v_B)(t^{\text{end}} - \Delta T)}{v_f} \end{aligned} \quad (23)$$

Note, by the way, that the total delay is not equal to area in  $B$  since it depends on the speeds in  $B$ , see equation (10). However, it is proportional to area  $B$  given the speed of  $B$ .

All variables in equation (23) can be substituted into basic variables, using the relationships stated in equations 12, 13 and 5. This yields:

$$D = \frac{1}{2} \frac{C_2 \Delta T^2 (r - 1)(rC_2 - \psi Q)}{C_2 - \psi Q} \quad (24)$$

Note that this is the same function as Olmstead (2) finds, equation (1). However, the assumptions we made are less restricting than his assumptions. Contrary to Olmstead, the traffic jam has a non-zero length. Vehicles enter at the traffic jam when they reach the tail of the queue, which in practice is upstream of the bottleneck location. Furthermore, the traffic jam dissolves from the head of the queue.

As expected (2), the delay is proportional to  $\Delta T^2$ . It also is a quadratic form  $r$ , but not proportional to the fraction of the capacity that remains,  $(1 - r)^2$ . In fact, it is proportional to  $(r - 1)(r - \frac{\psi Q}{C})$ . However, when the total capacity is used,  $\psi Q = C_2$ , then  $(\psi Q - rC_2)$  is proportional to  $1 - r$  and therefore the delay is proportional to  $(1 - r)^2$ .

For the test case given in section 4, the shape of the function can be seen in figure 2c. For this figure, we assumed that no spillback occurs at all, or in other words,  $\Delta x$  is larger than the queue length spillback.

### 5.1.3 Average delay

The number of drivers encountering delay can be computed using equation 8. The number of drivers that encounter delay is:

$$N = \int_0^{t^{\text{end}}} q_{\text{queue}} dt \quad (25)$$

Substituting  $q_{\text{queue}}$  (equation 7) and  $t^{\text{end}}$  (equation 16) now gives:

$$\begin{aligned} N &= \int_0^{t^{\text{end}}} k_A(v_A - \omega_{AB}) dt \\ &= k_A(v_A - \omega_{AB})t^{\text{end}} \end{aligned}$$

$$= k_A(v_A - \omega_{AB}) \frac{\omega_{BD}\Delta T}{\omega_{BD} - \omega_{AB}} \quad (26)$$

Note that this number of drivers is proportional to  $\Delta T$ . Because the total delay is proportional to  $\Delta T^2$ , the average delay per delayed driver is proportional to  $\Delta T$ :

$$\begin{aligned} A &= \frac{D}{N} = \frac{1}{2} \frac{C_2 \Delta T^2 (r-1)(rC_2 - \psi Q)}{C_2 - \psi Q} \bigg/ k_A(v_A - \omega_{AB}) \frac{\omega_{BD}\Delta T}{\omega_{BD} - \omega_{AB}} t^{\text{end}} \\ &= \Delta T \frac{1}{2} \frac{C_2 (r-1)(rC_2 - \psi Q)}{C_2 - \psi Q} \bigg/ k_A(v_A - \omega_{AB}) \frac{\omega_{BD}}{\omega_{BD} - \omega_{AB}} t^{\text{end}} \end{aligned} \quad (27)$$

## 5.2 Scenario 2: incident upstream of a junction

When an incident happens upstream of a junction, the junction can have an influence on the number of delayed travellers since also travellers to the other direction might be delayed. First, we discuss the traffic states that occur in such a case. Then, the total delay is determined, followed by an analysis which part of the delay is caused by the junction. Finally, the average delay per traveller will be computed. Like in scenario 1, back of the queue does not influence demand and thus the total delay could be calculated by vertical queuing model (section 3.4). In case the bottleneck would not be a fixed incident, but a moving bottleneck, the equations for the total delay would still hold. However, the average delay would be different. In this paper, we present the full dynamic extension of the queue because it provides insight into the dynamics because it is required to find the correct average delay, based on a fixed incident site.

### 5.2.1 Traffic states

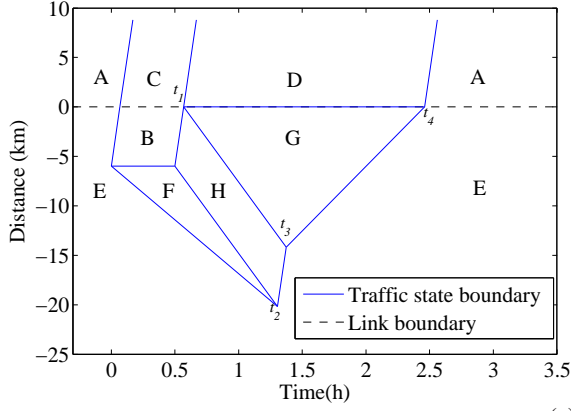
Suppose that an accident happens upstream of the junction (on link 1) which reduces the capacity temporarily to  $rC_1$ .

Figure 3a shows a pattern of a traffic situation that one would typically find (in this case,  $r = 0$  and  $\Delta T = 0.5$  and the incident takes place 6 km upstream of the bottleneck). This paragraph explains the traffic states; the states are also indicated in a flow-density plan in figure 3b. Areas A and E are the states in the non-incident situation, for links 2 and 1 respectively. There is an incident upstream of the bottleneck. Upstream of the incident a queue builds up (area F), and during the incident the outflow is lower (area B). This lower outflow reduces the demand to link 2 from  $t_0$  to  $t_1$ . After the capacity is restored, the traffic flows out of the traffic jam at the capacity of link 1 (area H). This flow is larger than the original demand. If the capacity of any of the downstream links is lower than the new demand to that link (the split fraction remains the same), a new area of congestion arises (area G). In this case, where  $\frac{\psi}{C_2} > \frac{1-\psi}{C_3}$ , link 2 forms a bottleneck in case  $\psi C_1 > C_2$ . Since the demand is high, the flow on the downstream link equals capacity, and on link 1 the flow is maximized by the outflow to the downstream links. In an equation, we could write

$$q_G = \min \left\{ \min \left( \frac{C_i}{\Psi_i} \right), C_1 \right\} \quad (28)$$

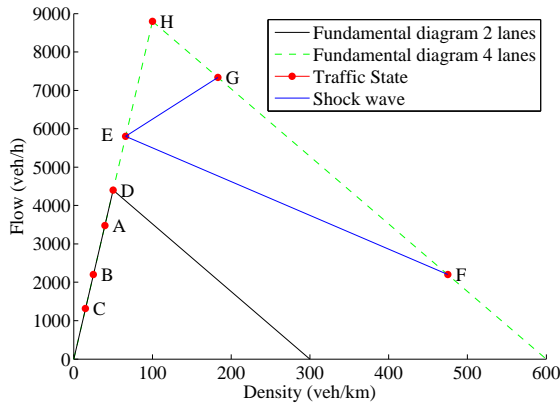
In the second case (equation 28, the demand is lower than the possible flows on the downstream links and the outflow of link 1 is not restricted by any of the downstream links. In that case, the problem reduces to the situation described in section 5.1. If the outflow is restricted, the traffic states as introduced in this section are applicable. Therefore, the traffic states as shown in figure 3 are only applicable if

$$C_1 > \min \left( \frac{C_i}{\Psi_i} \right) \quad (29)$$

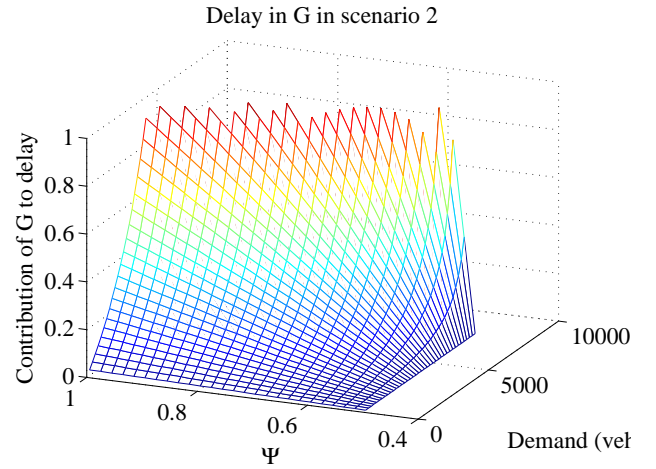


(a) Traffic states

State	Lanes	Flow	Congested
A	2	$\psi Q$	Free
B	4	$4rC$	Free
C	2	$4rC\psi$	Free
D	2	$2C$	Free
E	4	$Q$	Free
F	4	$4rC$	Congested
G	4	$\frac{2C}{\psi}$	Congested
H	4	$4C$	Free



(b) The traffic states on the density-flow diagram



(c) The contribution of the total delay that is encountered in G

**FIGURE 3 The traffic situation for an incident upstream of the junction**

### 5.2.2 Total delay

Only in areas  $F$  and  $G$  (see figure 3) travelers encounter delay. Using shockwave theory one can compute the total delay. Again, equation (11) has to be used to compute the total delay:

$$D = \int N_A(t) \frac{v_f - v_A}{v_f} dt \quad (30)$$

This formula is written for the general traffic state  $A$  ( $N_A, v_A$ ) which should be replaced by state  $F$  and  $G$  (i.e., use  $N_F, N_G, v_F, v_G$ ).

In this formula, the number of vehicles can be substituted by the length of the queue times the density:

$$N = kL(t) \quad (31)$$

$k$  can be derived using equation (5), whereas  $L$ , the distance between the boundaries of the traffic states can be computed using equation (6). These substitutions can be performed by a computer program. The result of substituting all variables in equation (31) gives the total delay in a closed equation:

$$D = \frac{1}{2} \frac{\Delta T^2 (r^2 \psi C_1^2 - C_1 \psi r Q - r C_1 C_2 + C_2 Q)}{C_2 - \psi Q} \quad (32)$$

Note that, like in the case without the influence of a junction, the delay is proportional to  $\Delta T^2$ .

### 5.2.3 Delay caused by junction

We will now show the impact the delay caused by an insufficient downstream capacity. In particular, we show the part of the delay that is encountered in  $G$  as function of  $\psi$ . Note that there is a restriction on  $\psi$ . We return to the test case (section 4) and assume that  $\psi$  is larger than or equal to 0.5. Since the situation is symmetrical, this assumption is not restrictive because either  $\psi$  or  $(1 - \psi)$  will be. Note that the delay in  $G$  is determined by the queue at moment  $t_2$  and the flow in  $G$  (see equation (28)).

In general, the initial network can only be congestion free with the following condition:

$$\psi_i Q < C_i \quad (33)$$

for all outflow links  $i$ . In our situation where link 2 forms the bottleneck (by convention of numbering)  $\psi$  varies between 0.5 and  $\frac{C_2}{Q}$ . For all demands between 0 and  $C_1$  and for all appropriate split fractions, we plot the fraction of delay encountered in  $G$  in figure 3c.

If the outflow rate ( $C_2/\psi$ ) equals the inflow rate, the queue length does not change, and thus the queue will not solve at all. This means that all vehicles that will pass, encounter delay. In the limit that  $\psi = \frac{C_2}{Q}$ , the queue will never solve. Therefore, all vehicles will encounter the delay and all other delays are negligible.

The queue length can also be described. Figure 3b shows that the length of the traffic jam does not decrease as  $\frac{C_2}{\psi}$  approaches the demand  $Q$ , the gradient of the line  $E - G$  decreases to 0. This means that the shock wave speed that defines the tail of the traffic jam approaches to zero, and the queue length remains the same.

### 5.2.4 Average delay

Let's finally consider the average delay per driver encountering congestion. Only in areas  $F$  and  $G$  (see figure 3b) there is a speed reduction. The propagation speed of the state boundaries  $BH$  and  $EH$  (figure 3b),  $\omega_{BH}$  and  $\omega_{EH}$  equals the free flow speed  $v_f$  and therefore the speed of the vehicles in area  $H$ . That means that the vehicles that enter area  $G$  via boundary  $HG$  are the same vehicles as exit area  $F$ . Therefore, the total number of delayed vehicles,  $N$ , is the sum of the vehicles encountering delay in area  $F$  ( $N_F$ ) and the number of vehicles that enter area  $F$  via boundary  $EG$  ( $N_{EG}$ ):

$$N = N_F + N_{EG} \quad (34)$$

Furthermore, we know

$$t_3 - t_1 = t_2 - \Delta T \quad (35)$$

For the number of drivers that face congestion in area  $F$ , the same holds as for the number of drivers facing congestion in scenario 1 (equation 26):

$$\begin{aligned} N_F &= \int_0^{t^{\text{end}}} k_E(v_E - \omega_{EF}) dt \\ &= k_E(v_E - \omega_{EF}) t^{\text{end}} \\ &= k_E(v_E - \omega_{EF}) \frac{\omega_{FH} \Delta T}{\omega_{FH} - \omega_{EF}} \end{aligned} \quad (36)$$

To compute  $N_{EG}$  we use equation (8):

$$N_{EG} = \int k_E(v_E - \omega_{EG}) dt \quad (37)$$

$$= (t_4 - t_3)k_E(v_E - \omega_{EG}) \quad (38)$$

We will now show that  $(t_4 - t_3)$  is proportional to the duration  $\Delta T$  to show that  $N_{EG}$  is proportional to  $\Delta T$ . Since  $FH$  is parallel to  $HG$  and  $BH$  is parallel to  $EH$ , we can write

$$\begin{aligned} t_3 &= t_1 + t_2 - \text{duration} \\ &= t_1 - \frac{\omega_{EF} * \Delta T}{\omega_{EF} - \omega_{FH}} \end{aligned} \quad (39)$$

The duration of the growing of area G,  $t_3 - t_1$  therefore is proportional to the duration  $\Delta T$ . This means that the maximum length of queuing area G is

$$\begin{aligned} x_3 &= \omega_{GH} * (t_3 - t_1) \\ &= -\omega_{GH} \frac{\omega_{EF} * \Delta T}{\omega_{EF} - \omega_{FH}} \end{aligned} \quad (40)$$

This length decreases with  $\omega_{EG}$  per unit time, so

$$t_4 - t_3 = x_3 / \omega_{EG} \quad (41)$$

Substituting equation (40) into equation 41 gives

$$t_4 - t_3 = \frac{-\omega_{GH} \omega_{EF} * \Delta T}{\omega_{EG} \omega_{EF} - \omega_{FH}} \quad (42)$$

This shows that  $t_4 - t_3$  is proportional to the duration and therefore  $N_{EG}$  is proportional to the duration.

We can now compute the total number of delayed vehicles, substituting equations (36), (38) and (42) in equation (34):

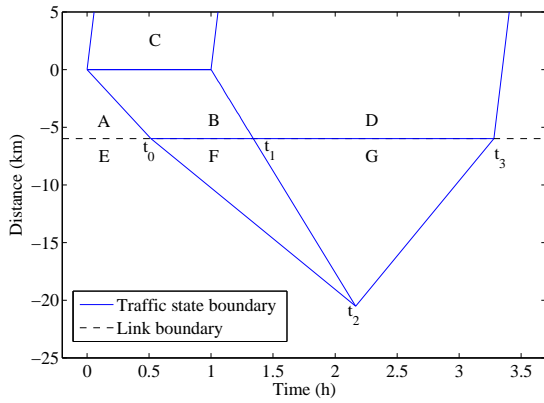
$$\begin{aligned} N &= N_F + N_{EG} \\ &= \frac{k_E(v_E - \omega_{EF})\omega_{FH}\Delta T}{\omega_{FH} - \omega_{EF}} + k_E(v_E - \omega_{EG}) \frac{-\omega_{GH} \omega_{EF} * \Delta T}{\omega_{EG} \omega_{EF} - \omega_{FH}} \end{aligned} \quad (43)$$

Note that the number of travelers that encounters delay is proportional to the duration  $\Delta T$ . Therefore, the average delay per traveler  $A$ , the total delay (equation 32) divided by the number of delayed travellers, is also proportional to the duration:

$$\begin{aligned} A &= D/N \\ &= \Delta T \left( \frac{1}{2} \frac{(r^2\psi C_1^2 - C_1\psi rQ - rC_I C_2 + C_{I2}Q)}{C_2 - \psi Q} / \dots \right. \\ &\quad \left. \frac{k_E(v_E - \omega_{EF})\omega_{FH}}{\omega_{FH} - \omega_{EF}} + k_E(v_E - \omega_{EG}) \frac{-\omega_{GH} \omega_{EF} * \Delta T}{\omega_{EG} \omega_{EF} - \omega_{FH}} \right) \end{aligned} \quad (44)$$

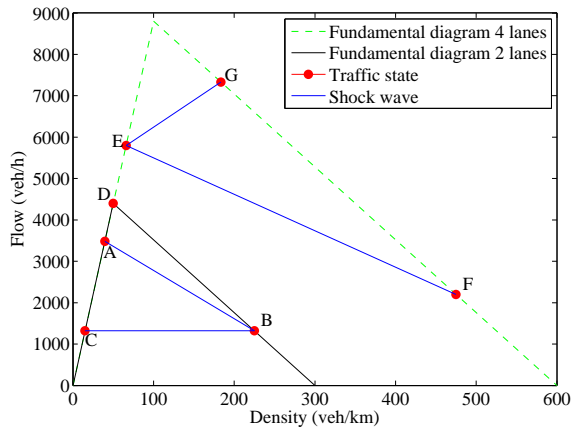
### 5.3 Scenario 3: queues longer than the distance to the junction

When an incident happens *downstream* of a junction, on link 2, the traffic states that occur depend on the length of the queue. If the queue is shorter than the distance to the upstream intersection and there is no restricting junction downstream, the traffic follows the pattern described in section 5.1. However, when it reaches the upstream junction, different traffic states occur. This section will first discuss the traffic states and the delays, and subsequently the influence of the junction.

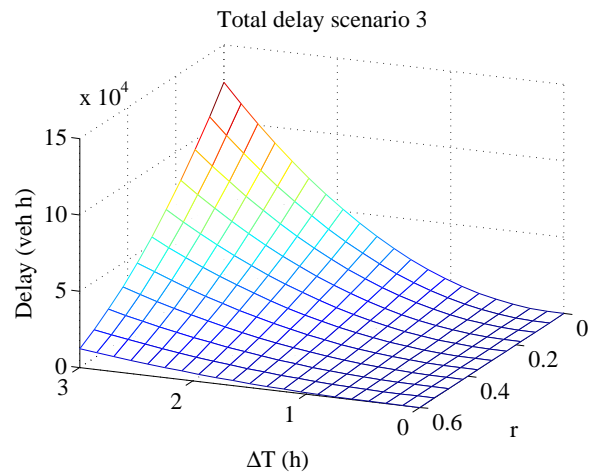


State	Lanes	Flow	Congested
A	2	$\psi Q$	Free
B	2	$2rC$	Congested
C	2	$2rC$	Free
D	2	$2C$	Free
E	4	$Q$	Free
F	4	$\frac{2rC}{\psi}$	Congested
G	4	$\frac{2C}{\psi}$	Congested

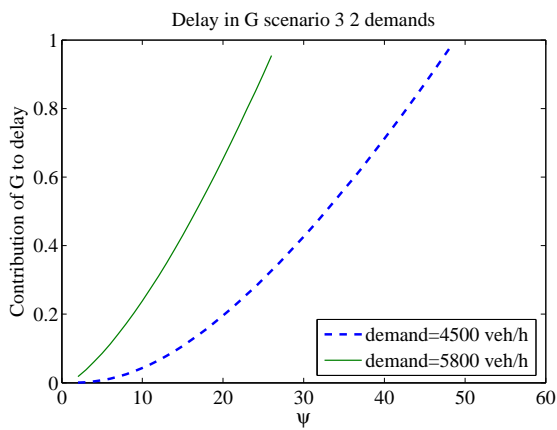
(a) Traffic states when the queue is longer than  $\Delta x$



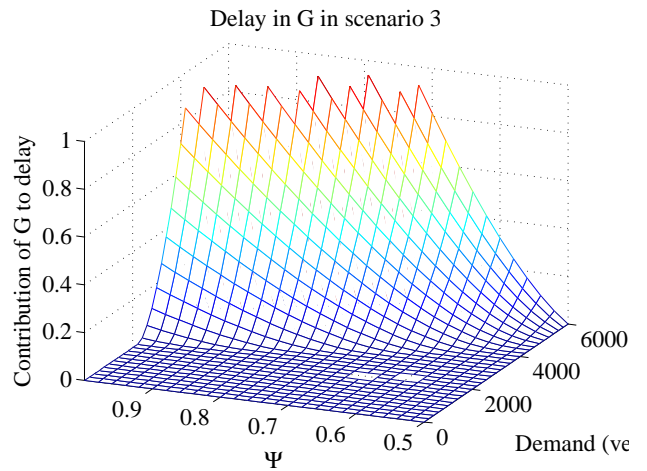
(b) Traffic states



(c) The delay caused by the incident



(d) The fraction of the delay in G as function of the split fraction. ( $r=0.3$  and  $\Delta T = 1$  hour)



(e) The fraction of the delay in G

**FIGURE 4** Queue longer than  $\Delta x$

### 5.3.1 Traffic states and delays

A typical pattern is given in Figure 4. For this particular graph, we use  $r = 0.3$ ,  $\Delta x = 6km$ , and  $\Delta T = 1h$ .

The congestion that the incident causes, spills back to the more upstream link. From the moment that congestion reaches the junction, the outflow of that link is reduced. In particular, fraction  $\psi$  that would like to turn to link 2 is reduced to  $rC_2$ . Consequently, the flow on link 1 is  $\frac{rC_2}{\psi}$ .

After the congestion on link 2 has solved (at  $t_1$ ) there is still congestion on link 1 if the turn fraction is not 50% – similar to the congestion described in section 5.2). The demand to links 2 and 3 equals the capacity of the link 1, rather than the demand. In the example case, the new demand from link 1 is  $4 * 2200 veh/h = 8800 veh/h$ . Link 2 cannot accommodate 60% of 8800 vehicles. Instead, the maximum capacity is 2 lanes times  $2200 veh/h/lane$ , hence there is still congestion on link 1. This means that the flow on link 1 is at maximum  $\frac{2*2200}{0.6}$ . Depending on the split fraction  $\psi$ , the flow can be restricted by either link 2 or link 3. A general prescription of the flow in area G can be found in equation 28. For the layout of our case study, this reduces to:

$$q_G = \min \left\{ \begin{array}{l} \frac{2C_2}{\psi}, \frac{2C_3}{1-\psi} \\ C_1 \end{array} \right\} \quad (45)$$

The traffic states are separated by boundaries. The speed at which these boundaries propagate can be computed using equation (6), the flow values as stated in figure 3a and the flow-speed relation in equation (5). Figure 3b shows the shock wave speeds.

The shock wave speeds determine the size of areas  $B$ ,  $F$  and  $G$  in figure 3a. The delay in these areas can be computed using equation (11). Although it is just a simple substitution of variables, the end result is quite long. The resulting delay is expressed as function of the duration and the capacity reduction of the incident. A plot of the delay function is shown in figure 3c.

### 5.3.2 Influence of junction

In figure 3e the influence of G is plotted as part of the total delay, just like in section 5.2. We had to fix the incident capacity and the duration; for the graph we choose  $r = 0.3$  and  $\Delta T = 1h$ . We also needed to fix the demand. In the figure we plot two lines: one for a demand of  $4000 veh/h$  and one for a demand of  $5800 veh/h$ . The first one, due to the restrictions on  $\psi$ , has a wider range of possible values for  $\psi$  that are possible compared to the demand of  $5800 veh/h$ . The same plot for the demands of  $4500 veh/h$  and  $5800 veh/h$  are similar of shape: at  $\psi = 0.5$  the contribution of G to the total demand is 0 and it raises (in a convex shape) to 1 (i.e., the full delay is encountered in G) for  $\psi = \frac{2C}{Q}$ .

Figure 3e shows the same fraction as function of the split fraction and the demand. This fraction raises to 1 in case that the demand approaches the critical demand. The reasoning for increase of this fraction is the same as for scenario 2, explained in section 5.2.

## 6 MATHEMATICAL IMPLICATIONS

The previous section described the delay caused by an incident. In this section we will rework these formulae to find closed-form expressions for the total delay and the average delay.

### 6.1 Total delay

Olmstead (2) derives an equation for the expected delay (equation 2) assuming point queues. We will now derive an expression for the expected value of the delay (or any other attribute  $Y$ ) in case it is proportional to the square of the duration (or any other property  $x$ ).



So suppose:

$$Y = cx^2 \quad (46)$$

In this equation,  $c$  is a constant. The variance of  $x$ ,  $\text{Var}(x)$ , is:

$$\text{Var}(x) = \langle x^2 \rangle - \langle x \rangle^2 \quad (47)$$

In these equations, the angle brackets mean the expected value. The expected value of  $Y$  can be written as:

$$\langle Y \rangle = \langle cx^2 \rangle = c \langle x^2 \rangle \quad (48)$$

Combining equations (48) and (47) gives:

$$\langle Y \rangle = c \langle x \rangle^2 + c \text{Var}(x) \quad (49)$$

This shows that if  $Y$  is proportional to the square of  $x$  the expected value of attribute  $Y$  can be derived from the expected value of  $x$  and the variance of  $x$ .

In scenarios 1 and 2, the delay is proportional to the square of the duration. Therefore, we can formulate an equation for the expected delay as a result of an incident with a stochastic delay, using the above theorem. We can now apply this result on the incident delays. In the case that there are no junctions, the total delay is proportional to the square of the delay (equation 24). This means that if assumption of vertical queues by Olmstead (2) is relaxed and we introduce realistic queuing, equation (2) is valid. Even more surprisingly, also in case of a junction downstream of the incident, the total delay is proportional to the square of the duration (equation 32). In case spillback occurs, this condition does not hold any more. We therefore can give an expression for the expected delay in case of no spillback:

$$\langle D \rangle = c \langle \Delta T \rangle^2 + c \text{Var}(\Delta T) = D(\langle \Delta T \rangle) + c \text{Var} \Delta T \quad (50)$$

with

$$\begin{aligned} c &= \frac{1}{2} \frac{C_2(r-1)(rC_2-\psi Q)}{C_2-\psi Q} && \text{no influence of a junction} \\ c &= \frac{(r^2\psi * C_1^2 - C_1\psi rQ - rC_1C_2 + C_2Q)}{C_2 - \psi Q} && \text{influence of downstream junction} \end{aligned} \quad (51)$$

Note that these equations are derived for the case that a traffic jam occupies space (so not only for point queues). However, the resulting formulation in case there is no influence of a junction is the same as Olmstead (2) derives (equation (2)).

## 6.2 Average delay

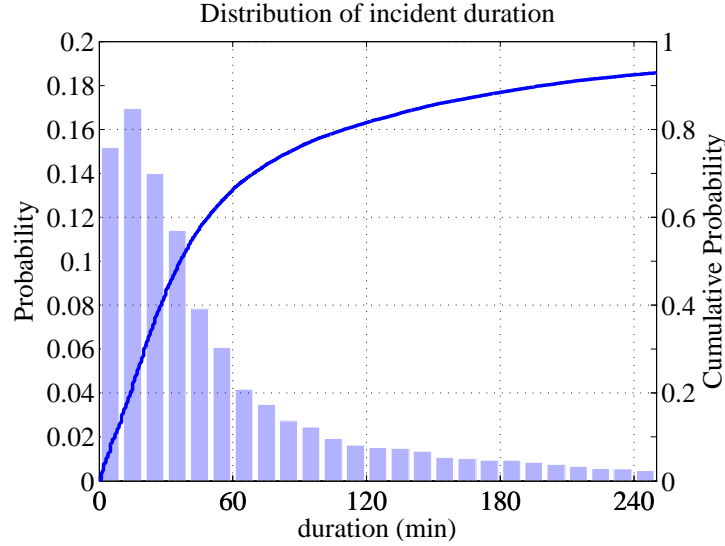
If there is no spillback (scenarios 1 and 2), the average delay per driver that encounters congestion is proportional to the duration (equations 27 for scenario 1 and 44 for scenario 2). We can now relate the variance of the delay per (delayed) driver to the variance of the duration. First we start by writing down an equation for the variance of the delay:

$$\text{Var}(A) = \langle A^2 \rangle - \langle A \rangle^2 \quad (52)$$

We now substitute  $A$  by  $c\Delta T$ , in which  $c$  is the proportionality constant expressed in equation (27) or (44):

$$\begin{aligned} \text{Var}(A) &= \langle (c\Delta T)^2 \rangle - \langle c\Delta T \rangle^2 \\ &= c^2 \langle (\Delta T)^2 \rangle - \langle \Delta T \rangle^2 = c^2 \text{Var}(\Delta T) \end{aligned} \quad (53)$$

This equation gives an expression for the variance of the delay for individual drivers.



**FIGURE 5** The duration of incidents

However, none of these simple rules holds for situations where spillback comes into play. Areas  $F$  and  $G$  (figure 4) can be seen as the congestion states for a special case of scenario 2, with a duration of  $(t_1 - t_0)^2$  and a distance to the junction of 0. The delay in area  $F$  and  $G$  therefore scales with  $(t_1 - t_0)^2$ , which can be computed using the following equations:

$$t_0 = \frac{\Delta x}{-\omega_{EF}} \quad (54)$$

$$t_1 = \frac{\Delta x}{-\omega_{FG}} \quad (55)$$

## 7 REAL-LIFE INCIDENT DATA

To show the impact of the equations, we analyse the duration of real-life incidents. We use a database of all incidents in the Netherlands between 1 January 2007 and 1 September 2007 provided by the Dutch Road Authority. The database consists of 55,176 freeway incidents; for 51,050 incidents there is a valid incident duration of which the distribution (cumulative distribution function and a histogram) is plotted in figure 5. It is a non-symmetrical distribution with a mean of 77 minutes and a standard deviation of 105 minutes. The median (36 minutes) is much smaller than the mean. Around 65% of the incidents is cleared in less than 60 minutes. There is also a long tail in the distribution: there are some incidents which take a long time to remove. The total delay of an incident with a duration of the average duration underestimates the expected value of total delay. According to equation (50), the correction that has to be made to this estimate is the variance, or the square of the standard deviation. The relative importance of the variance can be derived from the quotient between the square of the standard deviation and the square of the mean value. The value of this quotient is  $105^2/77^2 = 1.93$ , which means that using the average duration in assessing the total delay would give only  $1/1 + 1.93 = 34\%$  of the total expected delay.

Using the values for demand, capacity and the split fraction as mentioned in section 4 and assuming a reduction  $r$  of 50% due to the incident we can compute the delays using equation 50. This consist of 2 parts: one part in case all incidents had the same duration (the calculation one would make using the mean duration) and a part due to the stochasticity of the incident duration. This is a numerical evaluation of the concept presented in the previous paragraph. Note that if one would only use the mean duration in the

calculation of the total delay, one would only find around 1/3 of the total encountered delay. This fraction is based on the fact that the delay is quadratic and the distribution of the incident duration. For scenario 3, this fraction is even lower. How much depends on the specific conditions of the road layout and incident location compared to the junction, and the distribution of the incident duration.

## 8 CONCLUSIONS

This paper analyses traffic states occurring after an incident. Using shockwave theory, the traffic states that result from the incident are calculated. As opposed to a point queue model, the head and the tail of the queue are modeled separately and in this way the spatial extent of the queue is described properly. Using the traffic states, the delay is analytically calculated.

A formula for the total delay was derived. We found that in the scenarios without the tail of the queue reducing the flow to other links (the first two scenarios), the delay is proportional to the square of the duration of the blocking. Because this is not linear in the duration, the expected value of the delay is not the delay of the incident with the expected value of the duration. An expression for the expected delay as a function of the variance of the duration of the blocking was formulated. We also formulated the variance of the average delay per involved traveler as function of the variance of the delay. The proposed equations do not hold for moving bottlenecks: this is a topic for future research.

The delay in case that the queue spills back to an upstream link can be expressed analytically but the result is more complicated and cannot be captured in a simple equation. In that case there is no simple relationship between the expected delay caused by an incident having the expected delay and the expected delay caused by an incident with stochastic delay. We therefore conclude that to analyse the delays in a network, it is needed to analyse the consequences of incidents with various durations and, also in practice, the mean duration is not sufficient to analyse the delay.

## REFERENCES

- [1] Bates, J., J. Polak, P. Jones, and A. Cook, The valuation of reliability for personal travel. *Transportation Research Part E: Logistics and Transportation Review*, Vol. 37, No. 2-3, 2001, pp. 191–229.
- [2] Olmstead, T., Pitfall to Avoid when Estimating Incident-Induced Delay by Using Deterministic Queuing Models. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1683, 1999.
- [3] Li, J., C.-J. Lan, and X. Gu, Estimation of Incident Delay and Its Uncertainty on Freeway Networks. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1959, 2006.
- [4] Khattak, A., J. Schofer, and M. Wang, A simple time-sequential procedure for predicting freeway incident duration. *IVHS Journal*, Vol. 2, 1995, pp. 113–138.
- [5] Garib, A., A. E. Radwan, and H. Al-Deek, Estimating magnitude and duration of incident delays. *ASCE Journal of Transportation Engineering*, Vol. 123, 1997, pp. 459–466.
- [6] Knibbe, W. J. J., T. P. Alkim, J. F. W. Otten, and M. Y. Aidoo, Automated Estimation of Incident Duration on Dutch Highways. In *Proceedings of IEEE ITS Conference*, Toronto, Canada, 2006.
- [7] Wang, W., H. Chen, and M. C. Bell, Vehicle Breakdown Duration Modeling. *Journal of Transportation and Statistics*, Vol. 8, No. 1, 2005, pp. 75–84.

- [8] Boyles, S., D. Fajardo, and S. T. Waller, A Nave Bayesian Classifier for Incident Duration Prediction. In *Proceedings of the 86th Annual Meeting of the Transportation Research Board*, Washington D.C., 2007.
- [9] Zhang, N., B. Jiyang, L. Sun, S. P. Hoogendoorn, and S. Hoogendoorn-Lanser, Estimation of Incident Duration with a Combined Classification Tree and Regression Method. In *Proceedings of Sino Dutch cooperation on Intelligent Transport Systems and Road Safety*, Beijing, China, 2007.
- [10] Fu, L. and L. R. Rilett, Real-Time Estimation of Incident Delay in Dynamic and Stochastic Networks. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1603, 1997.
- [11] Vickrey, W. S., Congestion Theory and Transport Investment. *The American Economic Review*, Vol. 59, No. 2, 1969, pp. 251–260.
- [12] Newell, G. F., A simplified theory of kinematic waves in highway traffic, part II: Queueing at freeway bottlenecks. *Transportation Research Part B: Methodological*, Vol. 27, No. 4, 1993, pp. 289 – 303.
- [13] Daganzo, C. F., *Fundamentals of Transportation and Traffic Operations*. Pergamon, 1997.
- [14] Helbing, D., *Verkehrsdynamik*. Springer, Heidelberg, 1997.
- [15] Dutch Road Authority, *Nieuwe Ontwerprichtlijn Autosnelwegen*. Rijkswaterstaat Adviesdienst Verkeer en Vervoer, 2007.
- [16] Transportation Research Board, *Highway Capacity Manual*, 2000.