

**Assessment of the Impact of Incidents Near Bottlenecks:
Strategies to Reduce Delay**

Monica Menendez
(corresponding author)
ph.: 510.642.9907 or 510.965.1549
fax: 510.965.1549
e-mail: acinom76@yahoo.com
mailing address: 1207 Melville Sq. #312, Richmond, CA 94804

Carlos Daganzo
ph.: 510.642.3853
fax: 510.642.1246
e-mail: daganzo@ce.berkeley.edu
mailing address: 416A McLaughlin Hall #1720, Berkeley, CA 94720-1720

(March 15, 2003)

5736 words

ABSTRACT

This study evaluates how the location and duration of an incident affect delays near a recurrent bottleneck. Using conventional kinematic wave theory and some dimensional analysis, it provides the tools to determine whether an incident will cause “generalized delays” (i.e., delays that have a lingering effect for the whole length of the rush hour) according to its magnitude, location, and duration. The results apply to a broad range of cases, encompassing many types of facilities and incidents. Furthermore, the results can be used as a foundation for the development and implementation of new strategies to obtain significant reductions in delay. The value of fault-free surveillance is also analyzed and presented as part of an optimization problem for the location of roadside assistance vehicles. It is found that this value is very high; which could justify the installation of advanced traffic monitoring schemes near major bottlenecks.

1. INTRODUCTION

Traffic incidents are known for their disruptive consequences on traffic flow. If traffic is congested, they can affect large portions of a transportation network. Their effects can be magnified many fold if the incidents occur near major recurrent bottlenecks such as bridges or merges.

This paper uses kinematic wave theory (1, 2) to classify incidents occurring near recurrent bottlenecks according to the delay they add to the system. Formulas are presented to predict this extra delay as a function of the characteristics of the highway, the bottleneck and the incident—i.e., its magnitude, duration, and location. These formulas can be used to design strategies for the detection and removal of incidents with roadside assistance vehicles. This paper presents a simple case with a single link and a single bottleneck. Two strategies are compared: a system that integrates surveillance and response using only floating trucks (i.e. tow trucks moving around), and another that uses tow trucks only to remove incidents and a separate surveillance method (e.g., video cameras) to detect them. The results are surprising. It is found that separate surveillance reduces the number of tow trucks by more than 50%; sometimes considerably more. These results are important because they suggest how to best allocate freeway surveillance and assistance resources, and that the benefits of doing so can be considerable.

The paper is arranged as follows: Section 2, below, states the problem in terms of kinematic wave theory. Section 3 shows how a generic incident impact problem can be solved, and Section 4 summarizes the results in a readily usable way. Section 5 uses those results to find the optimal location of roadside assistance vehicles and the ensuing societal “cost”, assuming that incidents are instantaneously detected (e.g., with a video surveillance system). The Section also evaluates a system that uses floating trucks to both detect and remove incidents, and compares it with the video surveillance approach.

2. PROBLEM STATEMENT

This paper relates the consequences of an incident occurring near a bottleneck, to the characteristics of the incident, the bottleneck and the highway. Figures 1a and 1b show a schematic representation of an incident both upstream and downstream of a recurrent bottleneck.

The symbols appearing in the figure, Q and B , represent the highway capacity (veh/time), and the capacity of the recurrent bottleneck. The variable q represents the capacity at the incident site. It is assumed that the incident is located a distance d from the bottleneck, and that it creates immediately downstream a rubbernecking zone of length l where speed is reduced (drivers might reduce their speed because of safety reasons or just to look at the incident; the actual cause is not relevant to this paper). The capacity in the rubbernecking zone is q , even if the zone overlaps with the bottleneck, as shown in Figure 1c.

To model traffic conditions, the hydrodynamic theory of traffic flow is used (1, 2) with a triangular relationship between flow and density. Empirical evidence suggests that this is reasonable (3, 4, 5, 6). Figure 2 shows the flow-density diagrams for the three relevant portions of highway: the recurrent bottleneck, the incident site and the rest of the highway. Note that in the figure, the free-flow speed is assumed to be V everywhere, except in the rubbernecking zone, where it is $v \leq V$. We also assume, as shown in the figure, that the wave speed, W , is the same everywhere.

Since we are interested in examining the effect of incidents when they can do the most damage (i.e., during the rush hour) we will assume that the bottleneck is “active” when the

incident occurs. (An active bottleneck is defined (7) as one discharging at its maximum capacity with a queue upstream and free flowing conditions downstream). We will also limit the analysis to cases where $q < B$. This ensures meaningful results, for otherwise the incident would cause no queuing disturbance at all, given that it occurs downstream of the bottleneck or within the queue directly upstream of the bottleneck.

3. INCIDENT IMPACT MODEL

This section uses kinematic wave theory to show how the intrinsic characteristics of an incident (l, q, v) together with its location d and its duration, influence overall system delay. We shall see that if incidents are removed quickly enough, they cause delay to just a few vehicles. On the other hand, if their duration exceeds a “critical time”, t_c , the effects might last for the whole rush hour. This critical time turns out to play an important role in incident management strategies.

Figures 3a and 3b show four different scenarios in which an incident has occurred near a bottleneck. The time-space diagrams have been drawn according to the flow-density relations of Figure 2, and the different numbers correspond to the traffic states defined there. The dotted lines represent vehicle trajectories, the continuous light lines waves, and the dark lines the actual trajectories of typical vehicles with and without the incident.

Figure 3a shows two short-duration incidents, upstream and downstream of the bottleneck. In the upstream case (bottom of figure), the actual and the theoretical trajectories meet before reaching the bottleneck. Therefore, the typical vehicle suffers no delay at all, since it passes the bottleneck at the same time whether or not there is an incident. In the downstream case, vehicles that encounter the incident suffer some delay (localized delay), but the vehicles that come after the queue generated by the incident vanishes are not affected. Note too that in both cases of part (a) the flow through the bottleneck is the same with and without the incident.

Figure 3b shows the same two incidents, but with a longer duration. The critical difference is that in this case the waves reach the bottleneck, reducing its flow and delaying every vehicle from then on, even after the incident has been removed. Thus, in this case, the incident creates generalized delays that last for the length of the rush hour. Obviously, for a given incident location, we can define a critical duration t_c that separates cases (a) and (b). For this critical duration, the deceleration and acceleration waves meet right at the bottleneck.

Note that in Figure 3b the actual and theoretical trajectories for a typical vehicle never meet. Their final separation, τ , is the delay experienced by the vehicle, which (as the reader can verify) is the same for every vehicle until the end of the rush hour. Figure 4 shows an input-output diagram depicting the cumulative flows through the bottleneck for the bottom part of Figure 3b. (The top part of Figure 3b would yield a similar diagram). The left-most curve of Figure 4 gives the desired arrival time of each vehicle at the bottleneck, the middle curve the departure time without the incident, and the right-most curve the actual departure time. The horizontal separation between the last two curves, τ , is the extra delay caused by the incident. The diagram clearly shows that when a recurrent bottleneck is affected by an incident, its discharge rate decreases temporarily to q and the capacity loss is felt until the end of the rush hour. The shaded region in Figure 4 represents the total extra delay caused by the super-critical incident of Figure 3b. We now use these ideas to develop a general solution.

4. GENERAL SOLUTION

We consider the upstream case (bottom of Figure 3) first. As shown in Figure 3, the deceleration and acceleration waves change their speed if a rubbernecking zone exists. The acceleration wave

travels at a speed v in the rubbernecking zone, and at the free flow speed V further downstream. Consideration of the diagrams also reveals that the speeds of the deceleration shockwave in the rubbernecking zone (s) and in the rest of the highway (S) are:

$$s = \frac{B - q}{Q \left(\frac{1}{V} + \frac{1}{W} \right) - \frac{B}{W} - \frac{q}{v}}, \quad (1a)$$

and

$$S = \frac{B - q}{Q \left(\frac{1}{V} + \frac{1}{W} \right) - \frac{B}{W} - \frac{q}{V}}. \quad (1b)$$

Further analysis of Figure 3b reveals that the critical incident duration must satisfy:

$$\frac{t_c}{d} = \left(1 - \frac{l}{d} \right) \cdot \left[\frac{1}{S} - \frac{1}{V} \right] + \left(\frac{l}{d} \right) \cdot \left[\frac{1}{s} - \frac{1}{v} \right]. \quad (2a)$$

Note that if there are no rubbernecking effects ($v = V$), then $s = S$ and Equation 2a reduces to,

$$t_c = \left[\frac{d}{S} - \frac{d}{V} \right]. \quad (2b)$$

This is logical; the critical time is just the difference in the trip times of the acceleration and deceleration waves from the incident to the bottleneck. Since it is usually the case that $v \cong V$ and $V \gg S$, a rough approximate for t_c is

$$t_c \cong \frac{d}{S}. \quad (2c)$$

Since the general solution of our problem includes 8 input parameters (B, Q, q, V, v, W, l, d), plus the result, t_c , it is useful to express it in terms of as few dimensionless groups of variables as possible. We define $\alpha = l - q/B$, $0 \leq \alpha \leq 1$ (incident capacity as a function of bottleneck capacity), and $\beta = l - B/Q$, $0 \leq \beta \leq 1$ (bottleneck capacity as a function of highway capacity); and find after a few manipulations that Equations 1 and 2 reduce to:

$$\frac{t_c \cdot V}{d} = \frac{\left(\frac{\beta}{1 - \beta} \right) \left(\frac{V}{W} + 1 \right)}{\alpha} + \frac{l}{d} \cdot \frac{\left(1 - \frac{V}{v} \right)}{\alpha}. \quad (3a)$$

As in Equation 2a, the second term on the right-hand side can be neglected if, as it is usual, $v \cong V$ and $l \ll d$. Since $V/W \cong 5$ in all highways on which observations have been made (6, 8), we can then write:

$$\frac{t_c \cdot V}{d} = \frac{6}{(1 - \beta)} \cdot \frac{\beta}{\alpha}. \quad (3b)$$

Figure 5 displays Equation 3b.

If the analysis is now repeated for the downstream problem (top of Figure 3) we find that it is almost symmetric to that one upstream, with only two differences: in the downstream case (i) there is always some localized delay, and (ii) rubbernecking never affects t_c . Thus, Equation 3b and Figure 5 are better approximations when the incident occurs downstream of the bottleneck. The exact solution is then given by Equation 2b.

Furthermore, note that incidents occurring inside a bottleneck have exactly the same effect as incidents at a distance zero from the bottleneck. In that case the critical duration is zero

and the generalized delays cannot be avoided. Moreover, when an incident occurs inside the recurrent bottleneck, even if there is no lane blockage but only rubbernecking, there will be generalized delays. The same is true for an incident where l is greater than or equal to d , as in Figure 1c.

Finally note from Figure 4 that if an incident exceeds the critical duration by an amount Δ , then it induces an extra delay per vehicle equal to

$$\tau = \Delta \cdot \alpha . \quad (4)$$

This shows the importance of avoiding lengthy super-critical incidents with proper management strategies. As revealed in the next section, Equations 3 and 4 can be used to guide freeway service patrols.

5. APPLICATION: THE VALUE OF SURVEILLANCE

Our prior results are now applied to single link/single bottleneck systems to determine the optimal location of a roadside assistance vehicle, assuming that the freeway is under constant surveillance.

A link is a segment of highway (with one direction) with homogeneous traffic conditions between an on-ramp and an off-ramp. A network is composed of many links. Here, a link of variable length D , located directly upstream of the bottleneck, will be analyzed.

Recall that the extra delay per vehicle caused by an incident is described by Equation 4, and

$$\Delta = t_r + t_a - t_c , \quad (5)$$

where t_r is the removal time – how long the roadside assistance vehicle takes to remove the incident; t_a is the approach time – how long the roadside assistance vehicle takes to reach the location of the incident; and t_c is the critical time found in Section 4.

Since t_c is an increasing linear function of d (see Equation 2b), incidents close to the bottleneck are more likely to cause extra delays than incidents farther away. Figure 6 shows how τ depends on d for $t_a=0$ and $t_a>0$. The vertical separation between the two dark lines is the portion of the delay that can be avoided by faster response. Note that this separation declines toward zero for $d>Z$. If $d<Z$ it is impossible to avoid generalized delays, even with an instantaneous response. If $d>Z$ generalized delays can be avoided if t_a does not exceed a maximum $T_a(d)$ (dashed line in Figure 6). The distance Z is obtained by setting $t_r=t_c$ and using Equation 3b. The result is:

$$Z = \frac{t_r \cdot V \cdot \alpha \cdot (1 - \beta)}{6 \cdot \beta} . \quad (6)$$

Note that Equation 6 ignores rubbernecking effects.

The expression for $T_a(d)$ is:

$$T_a = 0 \quad \text{if } d < Z \quad (7a)$$

$$T_a = \frac{6 \cdot \beta}{\alpha \cdot (1 - \beta)} \cdot \frac{d}{V} - t_r . \quad \text{if } d > Z \quad (7b)$$

We now use Equations 5-7 to evaluate the effects of two different incident response strategies. With the first strategy (I), a single roadside assistance vehicle is stopped at a location to be determined, x distance units away from the bottleneck, and a separate surveillance system (e.g. video cameras) is used. With the second strategy (II), the assistance vehicle travels in a loop and acts as surveillance equipment. For both strategies we assume that the assistance vehicle travels at a speed u while on the link, and at a speed u' while outside (traveling to the upstream

end of the link to reposition itself) with a total cycle time C . We also assume that the assistance vehicle does not serve any other link while repositioning.

$$C = \left(\frac{D}{u} + \frac{D}{u'} \right). \quad (8)$$

Figures 7a and 7b show t_a as a function of the distance from the incident to the bottleneck for strategies I and II. The dashed line represents $T_a(d)$, the solid line in Figure 7a represents $t_a(d,x)$, and the point in Figure 7b represents $t_a \sim U[0,h]$. Avoidable delay γ only arises when $t_a > T_a(d)$; i.e., when the solid line or the point are above the dashed line. Note that the location of the incident, d , is uniformly distributed between 0 and D .

If strategy I is used (Figure 7a), the approach time depends on d , and the initial position of the assistance vehicle, x . Equations 9a and 9b can be used to calculate the expected avoidable delay per vehicle for two possible scenarios.

$$E(\gamma_I) = \alpha \cdot \left(\frac{1}{D} \right) \left(\int_0^D [t_a(d,x) - T_a(d)] dd \right) \quad \text{if } Z > D, \quad (9a)$$

and

$$E(\gamma_I) = \alpha \cdot \left(\frac{1}{D} \right) \left(\int_0^{Z_1} [t_a(d,x) - T_a(d)] dd \right) + \alpha \cdot \left(\frac{1}{D} \right) \left(\int_x^{Z_2} [t_a(d,x) - T_a(d)] dd \right) \quad \text{if } Z < D, \quad (9b)$$

where

$$Z_1 = \frac{x/u + t_r}{\left(1/u + 6 \cdot \frac{\beta}{\alpha \cdot V \cdot (1-\beta)} \right)}, \quad \text{and} \quad (10a)$$

$$Z_2 = \frac{x/u + t_r + C}{\left(1/u + 6 \cdot \frac{\beta}{\alpha \cdot V \cdot (1-\beta)} \right)}. \quad (10b)$$

If $Z > D$ (Equation 9a), the optimal position of the truck is at the end of the link, and no optimization is required ($x^* = D$). On the other hand, if $Z < D$ (Equation 9b), one needs to find the optimal value of x that minimizes $E(\gamma_I)$. (Fortunately, when $Z < D$, x^* is always greater than Z , and Equation 9b is always valid). The solution must be feasible ($Z < x < D$), otherwise, the optimal location of the truck is at the end of the link ($x^* = D$), and $E(\gamma_I)$ can be computed using only the first term on the right-hand side of Equation 9b.

If strategy II is used (Figure 7b) the expected approach time at every point is uniformly distributed with values between 0 and the headway h ,

$$h = \frac{C}{N} \quad (11)$$

where N is the number of roadside assistance vehicles ($N=1$ in Figure 7b) and C is the cycle time defined in Equation 8. Hence, the expected avoidable delay at a location d is,

$$E(\gamma_{II} | d, t_a, N) = \alpha \cdot \max\{t_a - T_a(d), 0\}. \quad (12)$$

Therefore, the expected avoidable delay across the whole link is:

$$E(\gamma_{II}) = \alpha \cdot \left(\frac{1}{D} \right) \left(\int_0^D \frac{h}{2} dd \right) \quad \text{if } D < Z, \quad (13a)$$

$$E(\gamma_{II}) = \alpha \cdot \left(\frac{1}{D} \right) \left(\int_0^Z \frac{h}{2} dd \right) + \alpha \cdot \left(\frac{1}{D} \right) \int_Z^D \left[\left(\frac{h - T_a(d)}{h} \right) \cdot \left(\frac{h - T_a(d)}{2} \right) \right] dd \quad \text{if } Z < D < Z_3, \quad (13b)$$

and

$$E(\gamma_{II}) = \alpha \cdot \left(\frac{1}{D} \right) \left(\int_0^Z \frac{h}{2} dd \right) + \alpha \cdot \left(\frac{1}{D} \right) \int_Z^{Z_3} \left[\left(\frac{h - T_a(d)}{h} \right) \cdot \left(\frac{h - T_a(d)}{2} \right) \right] dd \quad \text{if } Z < Z_3 < D, \quad (13c)$$

where

$$Z_3 = \frac{(t_r + h) \cdot V \cdot \alpha \cdot (1 - \beta)}{6 \cdot \beta}. \quad (14)$$

Finally, if one wants to know the value of permanent surveillance, one can solve for N making the expected delay for strategies I and II equal. The obtained value represents the number of roadside assistance vehicles required to obtain the same service with floating trucks than with one stopped truck. The magnitude of N will dictate the best strategy to be used. Such decision should be based on the cost of operating $N-1$ extra trucks versus the cost of implementing new surveillance technology.

Figure 8 shows $N-1$ as a function of the length of the link D ($0.25 \text{ miles } (0.4 \text{ km}) < D < 3 \text{ miles } (4.8 \text{ km})$). Some of the values have been fixed so numerical results can be provided.

$V = 60 \text{ mph } (96 \text{ km/hr})$,

$W = 12 \text{ mph } (19 \text{ km/hr})$,

$u = 25 \text{ mph } (40 \text{ km/hr})$, and

$u' = 15 \text{ mph } (24 \text{ km/hr})$.

The removal time t_r has been set to two different values: 5 min if the incident is removed quickly, and 20 min if the incident is removed slowly. Finally, the magnitude of the accident α has been set to three different levels: 0.75 for large incidents, 0.5 for medium size incidents, and 0.25 for small incidents. The graphs show a family of curves for different bottleneck sizes, $\beta = 0.1, 0.3, 0.5$, and 0.7 .

Each curve shows at most three regimes. The first one for $Z > x^* = D$, the second one for $Z < x^* < D$, and the third one for $Z < x^* = D$. In the first regime the curve is totally flat, while in the second and third regimes the curve bends up and down, respectively.

The results shown in these graphs are surprising. In most of the cases the $N-1$ values are above 1 . Moreover, when the capacity of the recurrent bottleneck is similar to that of the highway ($\beta = 0.1$), the $N-1$ values do not vary too much with the length of the link. In addition, the number of extra trucks is generally smaller for long removal times and/or more severe incidents (larger α). In strategy I the optimal location of the truck with respect to the bottleneck is affected by the magnitude of the incident (the smaller the incident, the closer to the bottleneck). However, in order to achieve the same results with strategy II, more trucks are required. Hence, strategy I becomes more valuable in cases where the magnitude of the incident is relatively small or its removal time is relatively short.

In summary, the graphs in Figure 8 show that very often a separate surveillance system reduces the number of roadside assistance vehicles by more than 50%. This suggests that a widespread used of a strategy I might bring substantial benefits. With strategy I, a smaller number of roadside assistance vehicles can be used to provide the same service as with strategy II, or if the same number of assistance vehicles is used, better service can be provided.

6. FINAL REMARKS

Nowadays, most of the Departments of Transportation recognize the significance of incidents near recurrent bottlenecks. The next step is to categorize those incidents according to their location, magnitude, and duration, and be able to respond accordingly. Allocation of resources as well as selection of the surveillance equipment and accidents' removal strategies can still be improved.

The example presented in Section 5 is a basic case illustrating one of many applications of the analysis performed in the first four sections. The results are very important because they suggest that considerable benefits can be obtained if a separate surveillance system is used in the area close to the bottleneck. These findings are consistent with Caltrans' current practices. California's Department of Transportation generally uses floating trucks to detect and remove traffic incidents. However, in locations close to recurrent bottlenecks such as the Bay Bridge in San Francisco, Caltrans employs a separate surveillance system and keeps its trucks stopped at one place.

This paper is the first step towards more complete optimization problems. The ideas here can be extended to more complex networks and more realistic problems including:

- a distribution of different sizes of incidents, probably based on historical data,
- incidents upstream of the queue,
- multiple incidents,
- multiple roadside assistance vehicles,
- networks of inhomogeneous links,
- bottlenecks with multiple approaches, and
- multiple bottlenecks.

ACKNOWLEDGMENT

This material is based upon work supported by a National Science Foundation Graduate Research Fellowship, a grant from the Transportation Center at the University of California, Berkeley, and the National Science Foundation Research Grant CMS- 0313317 to the University of California, Berkeley.

REFERENCES

1. Lighthill, M.J. and G.B. Whitham. On Kinematic Waves II, A Theory of Traffic Flow on Long Crowded Roads. *Proceedings of the Royal Society*, Vol. A 229, 1955, pp. 317-345.
2. Richards, P.I. Shock Waves on the Highway." *Operations Research*, Vol. 4, 1956, pp. 42-51.
3. Hall, F.L., B.L. Allen, and M.A. Gunter. Empirical Analysis of Freeway Flow-Density Relationships. *Transportation Research Part A*, Vol. 20A, No. 3, 1986, pp. 197-210.
4. Banks, J.H. Freeway Speed-Flow-Concentration Relationships: More Evidence and Interpretations. *Transportation Research Record 1225*, 1989, pp. 53-60.
5. Cassidy, M.J. Bivariate Relations in Nearly Stationary Highway Traffic. *Transportation Research Part B*, Vol. 32B, No. 1, 1998, pp. 49-59.

6. Windover, R.J. and M.J. Cassidy. Some Observed Details of Freeway Traffic Evolution. *Transportation Research Part A*, Vol. 35A, 2001, pp. 881-894.
7. Daganzo, C.F. *Fundamentals of Transportation and Traffic Operations*. Pergamon-Elsevier, Oxford, 1997.
8. Cassidy, M.J. and M. Mauch. An Observed Traffic Pattern in Long Freeway Queues. *Transportation Research Part A*, Vol. 35A, 2001, pp. 143-156.

LIST OF FIGURES

FIGURE 1 Schematic representation of an incident relative to a recurrent bottleneck.

FIGURE 2 Flow-Density diagrams.

FIGURE 3 Time-Space diagram for incidents occurring upstream and downstream of a recurrent bottleneck.

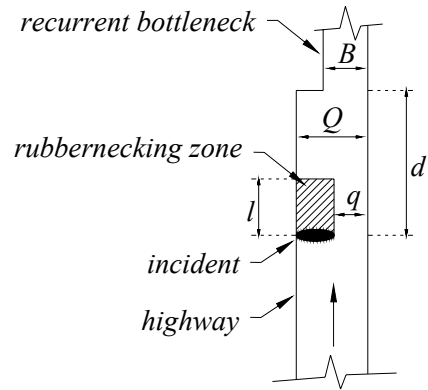
FIGURE 4 Input-Output diagram.

FIGURE 5 General solution for the critical time.

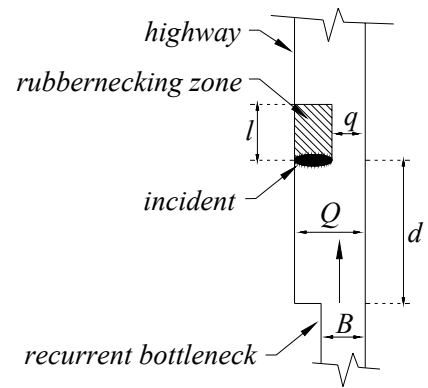
FIGURE 6 Extra delay as a function of distance.

FIGURE 7 Approach time as a function of distance.

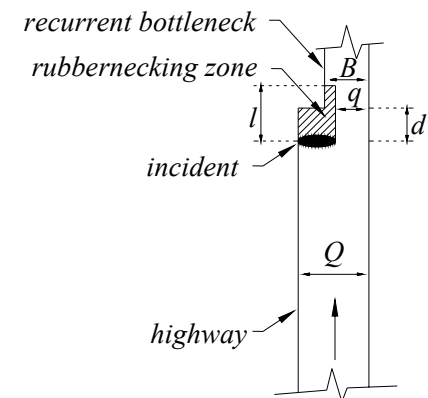
FIGURE 8 Value of surveillance.



(a)



(b)



(c)

FIGURE 1 Schematic representation of an incident
(a) upstream of a recurrent bottleneck,
(b) downstream of a recurrent bottleneck, and
(c) overlapping the recurrent bottleneck.

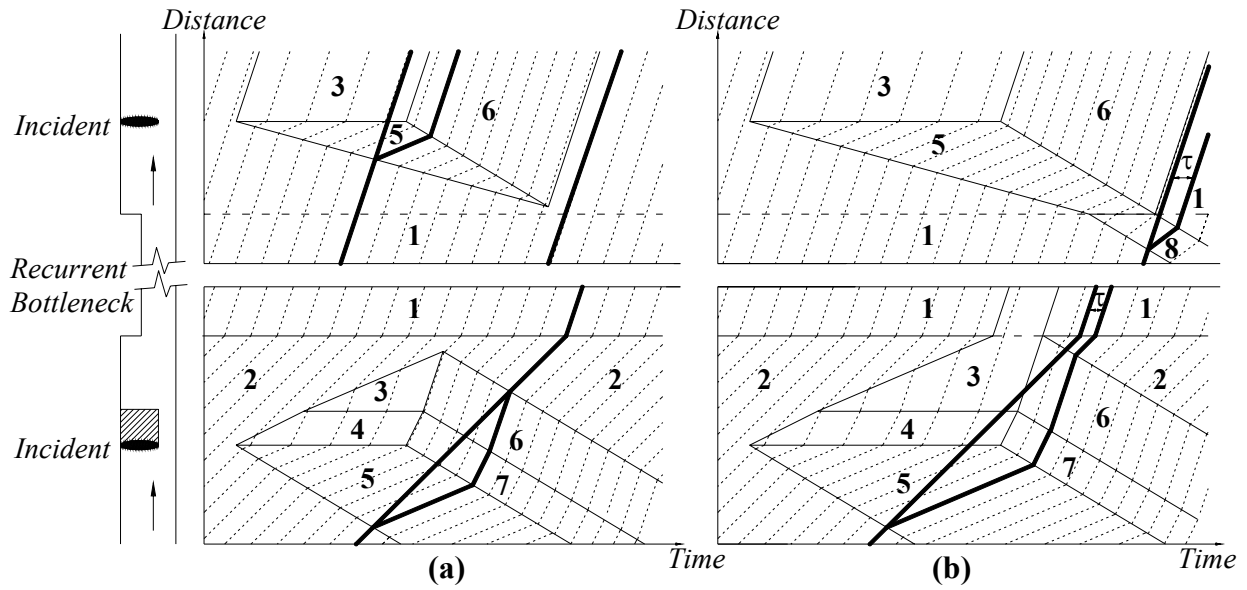


FIGURE 3 Time-Space diagram for incidents occurring upstream and downstream of a recurrent bottleneck with
(a) a short duration, and
(b) a super-critical duration.

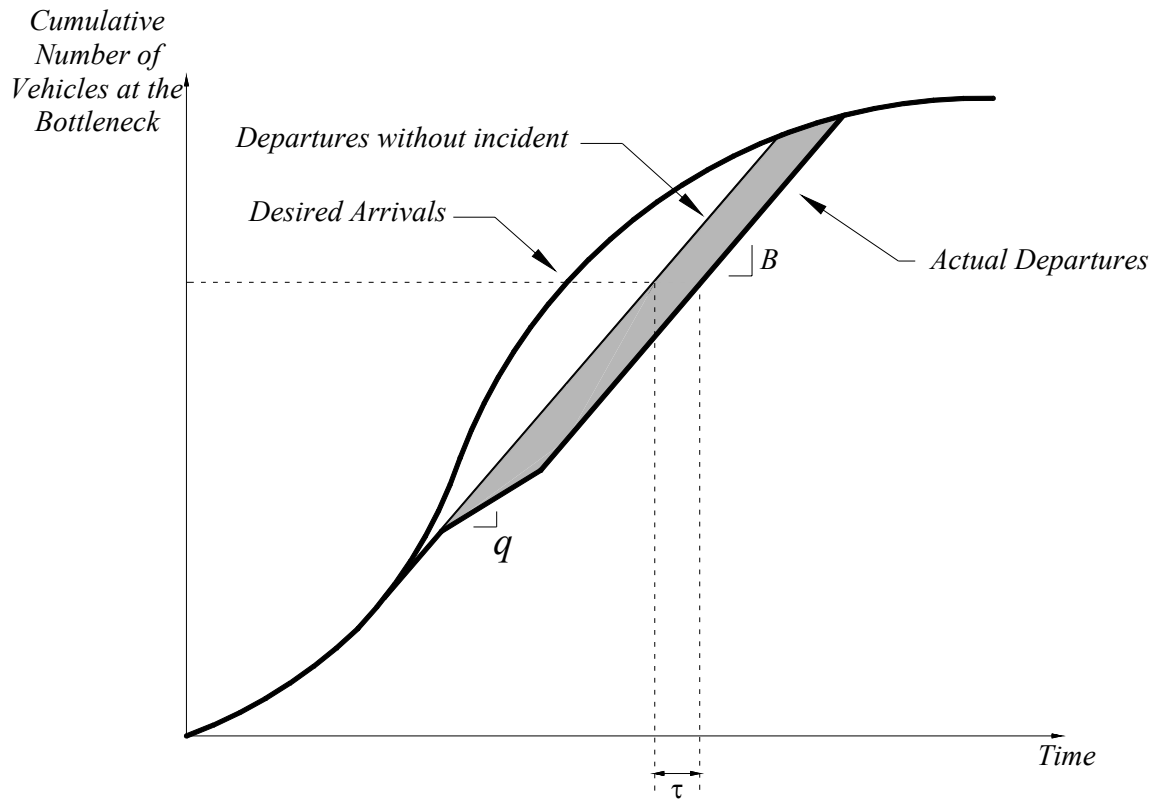


FIGURE 4 Input-Output diagram.

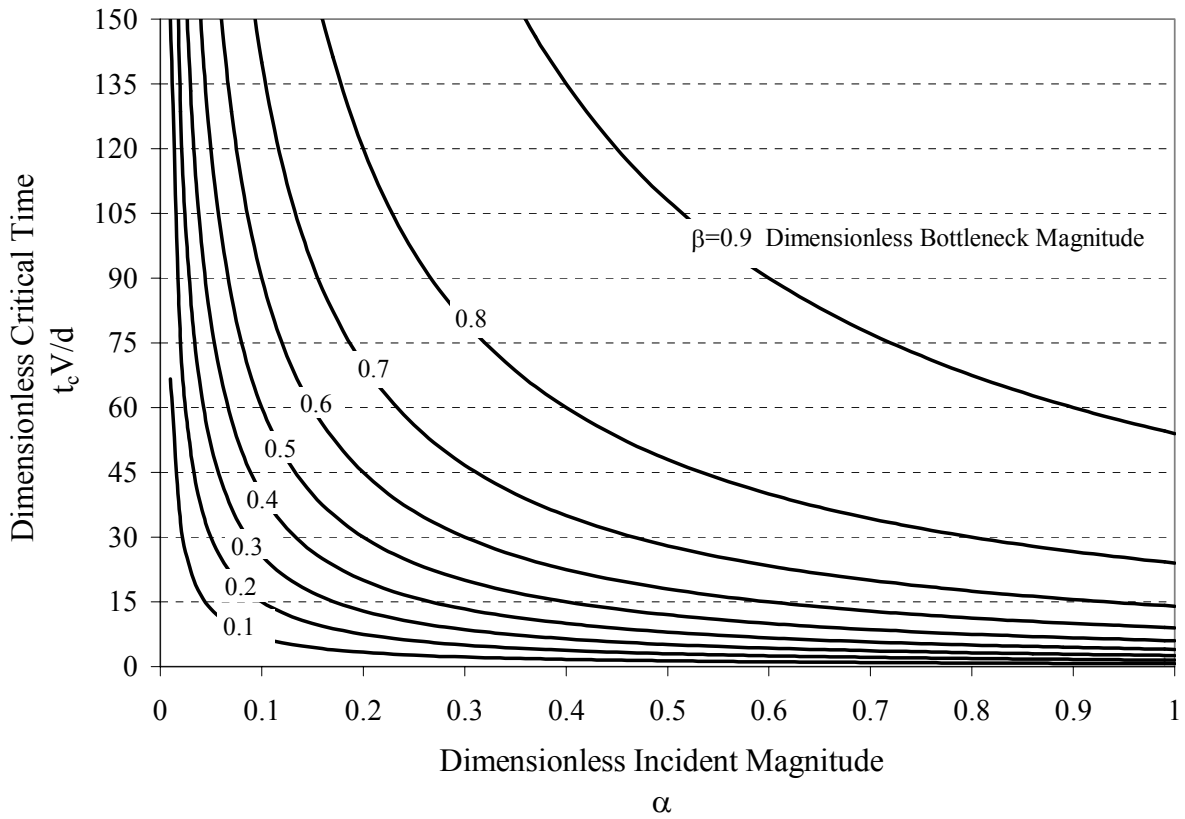


FIGURE 5 General solution for the critical time.

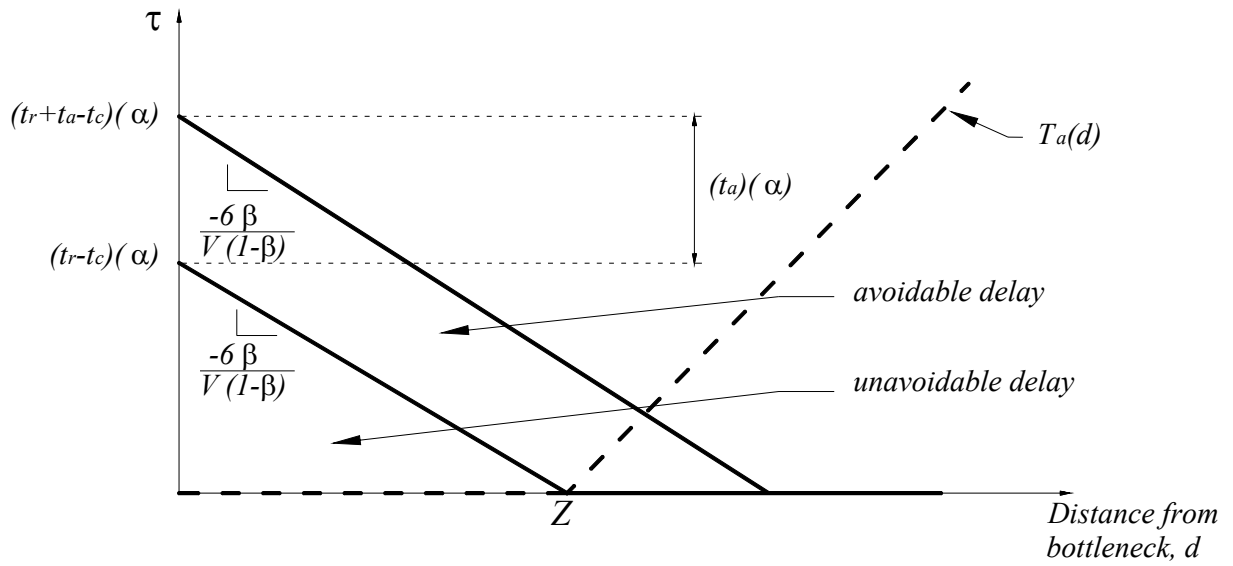


FIGURE 6 Extra delay as a function of distance.

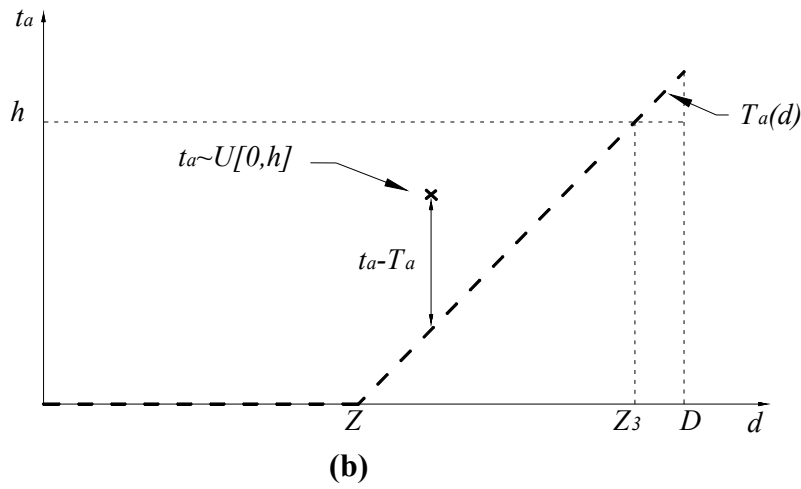
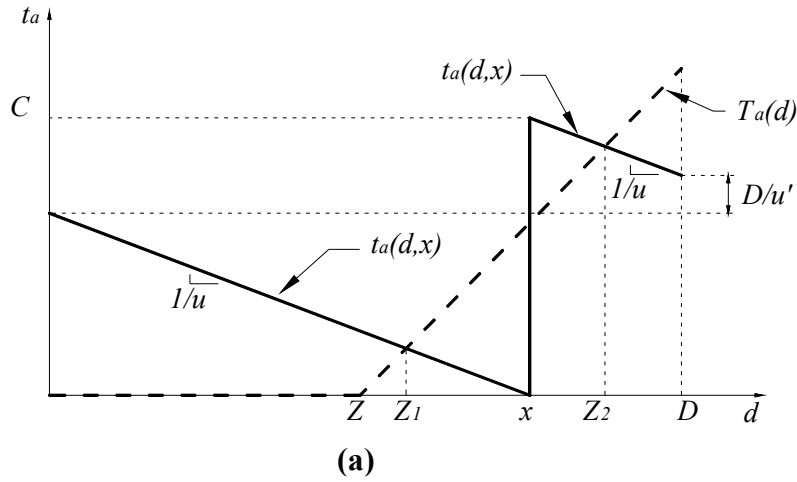


FIGURE 7 Approach time as a function of distance
 (a) strategy I, and
 (b) strategy II.

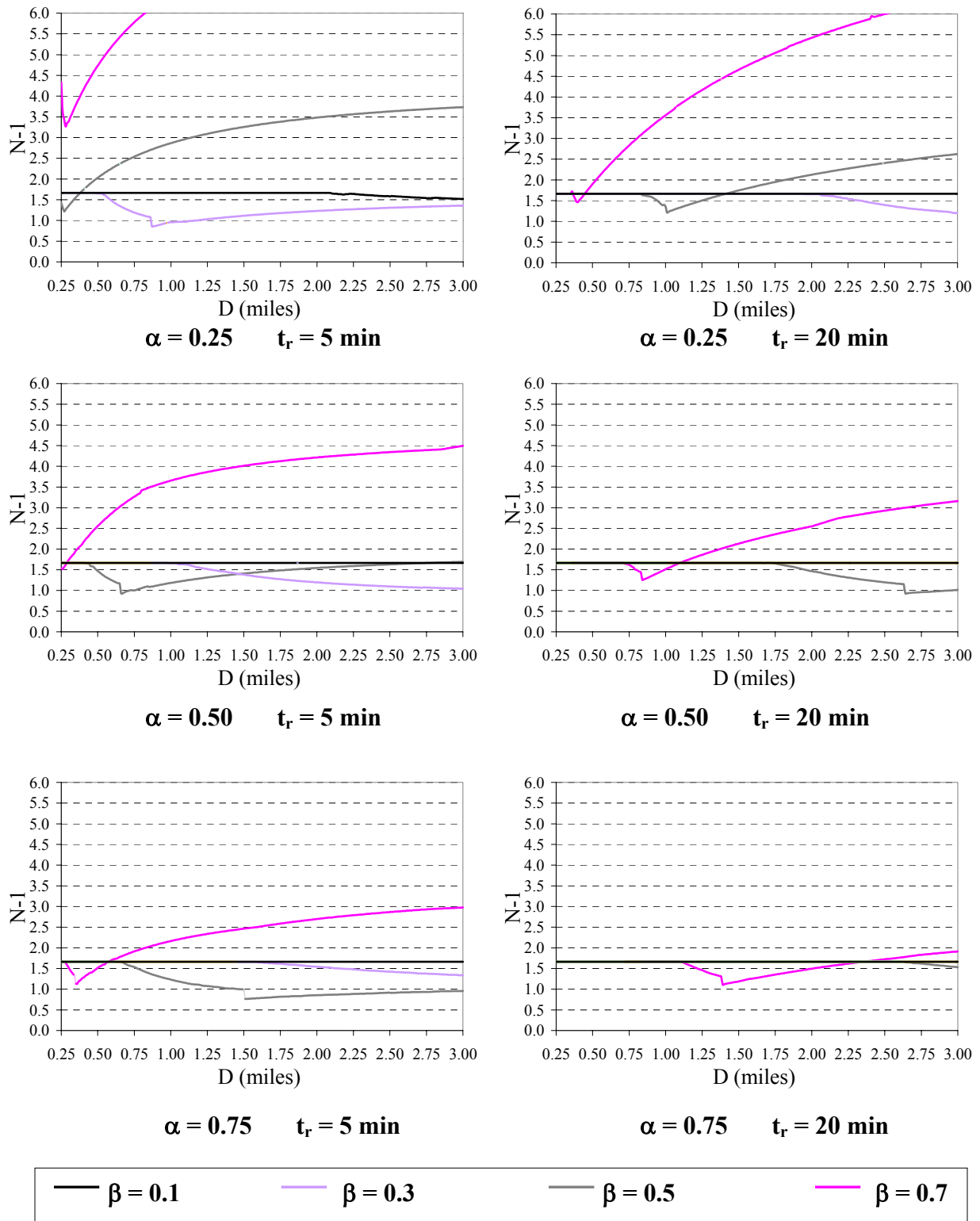


FIGURE 8 Value of surveillance. (1 mile = 1.61 km)