

# UC Berkeley

## Working Papers

### Title

Incident Dispatching, Clearance and Delay

### Permalink

<https://escholarship.org/uc/item/2pp689vn>

### Author

Hall, Randolph

### Publication Date

2000-08-01

**This paper has been mechanically scanned. Some errors may have been inadvertently introduced.**

CALIFORNIA PATH PROGRAM  
INSTITUTE OF TRANSPORTATION STUDIES  
UNIVERSITY OF CALIFORNIA, BERKELEY

# **Incident Dispatching, Clearance and Delay**

**Randolph Hall**

*University of Southern California*

**California PATH Working Paper  
UCB-ITS-PWP-2000-14**

This work was performed as part of the California PATH Program of the University of California, in cooperation with the State of California Business, Transportation, and Housing Agency, Department of Transportation; and the United States Department Transportation, Federal Highway Administration.

The contents of this report reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California. This report does not constitute a standard, specification, or regulation.

Report for MOU 354

August 2000

ISSN 1055-1417

# **INCIDENT DISPATCHING, CLEARANCE AND DELAY**

June 11,2000

Randolph W. Hall  
Dept. of Industrial and Systems Engineering  
University of Southern California  
Los Angeles, CA 90089-0193

## TABLE OF CONTENTS

	Abstract .....	ii
	Executive summary .....	iii
1.	Introduction .....	1
2.	Delay and the distribution of response time .....	4
3.	Response time .....	<b>9</b>
4.	Equilibration effects .....	26
5.	Summary .....	30
6.	References .....	32

## LIST OF FIGURES

1.	Breakeven value of $k_2$ as a function of $k_1$ and $C(W_1)$ .....	8
2.	Vehicle dispatching directions .....	12
3.	Coefficient of variation for individual beats .....	15
4.	Incident can be served by vehicle coming from any of four directions .....	17
5.	Expected response time for constantly spaced response vehicles as a function of entrance separation .....	19
6.	Squared coefficient of variation for response time for constantly spaced response vehicles as a function of entrance separation .....	20
7.	Expected response time for Poisson process response vehicles as a function of entrance separation; shown for $p$ values of 0, .1, .25 and .5. ....	22
8.	Expected response time with congestion delay. constant spacing; $p = .1$ or $.5$ . opposite queue = 0 or $.5$ x same side queue .....	24
9.	Squared coefficient of variation for response time with congestion delay. constant spacing; $p = .1$ or $.5$ .....	25
10.	Equilibrated utilization versus arrival rate ( $\rho$ ): $k = .4$ . $E(S) = 1.2.4$ and 6 .....	28
11.	Equilibrated service time ( $t$ ) as a function of arrival rate ( $\rho$ ): $k = .4$ . $E(S) = 1.2.4$ and 6 .....	29

## **ABSTRACT**

This report models response times and delays for highway incidents, accounting for spacing between interchanges and the time penalty for changing directions, enabling a response vehicle to reach an incident on the opposite side of the highway. A fundamental question in dispatching incident crews is whether to send the closest vehicle that is currently available or to wait for another vehicle to become available that is even closer. Waiting for a closer vehicle is advantageous because service time is effectively reduced, adding to capacity and providing stability at higher levels of utilization. But waiting for a vehicle to become available adds uncertainty, which contributes to expected traffic delay. As a consequence, any reasonably robust dispatch strategy must provide for a hybridization of the two objectives, trading-off greater certainty in response time against stability at higher utilization levels.

## EXECUTIVE SUMMARY

Incidents – such as collisions, stalls, and dropped loads -- are known to be a major source of highway delay. The amount of delay occurring during an incident depends on three primary factors: (1) the nature of the incidents, (2) roadway conditions, and (3) execution of incident clearance. Clearly, incidents that block more lanes and require more equipment to be cleared (e.g., those involving heavy duty trucks), will create more delay. The amount of delay will also increase when the roadway operates close to capacity and does not have alternatives for diverting traffic (either shoulders or parallel roadways). And delay also depends on how quickly the incident can be cleared and the actions taken during the incident to ensure smooth traffic flow.

The focus of this paper falls in the area of incident clearance. For the purposes of this paper, clearance time can be divided into four elements, which we call: (1) detection time, (2) dispatch time, (3) response time, and (4) service time. The specific focus of this paper is on the response time component of the dispatch time, and the contribution of dispatching policies to delay. Dispatch time merits special attention because it is one of the more controllable elements of clearance time. We consider here policies for dispatching mobile emergency crews, such as police officers and freeway-service-patrol trucks. The key characteristic is that the crews move around the network instead of residing at a stationary base (as is the case for fire crews).

A fundamental question in dispatching incident crews is whether to send the closest vehicle that is currently available or to wait for another vehicle to become available that is even closer. Waiting for a closer vehicle is advantageous because service time is effectively reduced, adding to capacity and providing stability at higher

levels of utilization. But waiting for a vehicle to become available adds uncertainty, which contributes to expected traffic delay. As a consequence, any reasonably robust dispatch strategy must provide for a hybridization of the two objectives, trading-off greater certainty in response time against stability at higher utilization levels.

In the case of individual beats, expected response time is a linear function of both the interchange spacing and the time penalty for changing direction of travel on the highway. On the other hand, with rolling beats (either with fixed spacing or Poisson process locations), increases in interchange spacing and the direction-change penalty do not cause expected response time to increase without bound. Instead, it approaches a limit, for which the responding vehicle always reaches the incident from an upstream location on the same side of the highway. However, rolling beats, in which the closest available vehicle is dispatched to the incident, have the drawback that they become unstable more easily.



# 1. INTRODUCTION

The amount of delay occurring during an incident depends on three primary factors: (1) the nature of the incidents, (2) roadway conditions, and (3) execution of incident clearance. Clearly, incidents that block more lanes and require more equipment to be cleared (e.g., those involving heavy duty trucks) create more delay. The delay will also increase when the roadway operates close to capacity and does not have alternatives for diverting traffic (either shoulders or parallel roadways). And delay also depends on how quickly the incident can be cleared, and the actions taken during the incident to ensure smooth traffic flow.

The incident clearance process can be divided into four elements: (1) detection, (2) dispatch, (3) response, and (4) service. Detection time is the time from when the incident occurs until the emergency response agency detects the presence of the incident. Dispatch time is the time from detection until the time that an emergency crew (or crews) is dispatched to the incident. Response time is the travel time for the emergency response crew to the scene of the incident. Last, service time is the time required to remove the incident and restore traffic once the emergency crew (or crews) has arrived at the scene.

The specific focus of this paper is on response and dispatch time, and the contribution of these to congestion and delay. There already exists a very large literature on incident detection, as well as many empirical studies on incident clearance times (Nam and Mannering, 2000, provide a recent review). Dispatching and response to highway incidents have received much less attention from an operational perspective. We consider here policies for dispatching mobile emergency crews, such as police officers

and freeway-service-patrol trucks. The key characteristic is that the crews move around the network instead of residing at a stationary base (as is the case for fire crews).

Dispatching processes and response times for police agencies is a well-studied topic in the field of operations research. Police are typically modeled as a spatial queueing system in which the servers (police cars) are mobile. The response time depends on the density of servers (cars per square mile), their overall utilization (ratio of demand to capacity) and the policy for dispatching officers. As the utilization increases, a greater percentage of cars are busy at any given time. Effectively, this causes the density of cars to decline and the response time to increase.

The most famous research in this area is the hypercube model developed by Larson (1974), along with related research by the author (e.g., Larson 1972; Larson and McKnew, 1982; Larson and Rich, 1987). The hypercube model is a stochastic queueing models that accounts for the assignment of patrol cars to districts (or beats), and rules for dispatching officers within and across districts. More recently, the spatial queueing approach has been extended to systems in which multiple cars must be dispatched (e.g., Green (1984), Green and Kolesar (1984), Green and Kolesar (1989), Ittimakin and Edward (1991)).

The design of emergency response systems has also been studied by a number of authors, including an extensive series of work through RAND in the 1970s (for example, Chaiken and Dormont, 1978; Ignall et al, 1978; Kolesar et al, 1975). One of the notable findings to come out of this work is the “square-root” rule for estimating average response distance to an incident. By this model, the response distance equals a constant multiplied by the square-root of  $1/\rho$ , where  $\rho$  is the density of patrol cars per unit area.

Ignall et al (1978) reported that this square-root relationship is even a reasonable approximation when the number of busy patrol cars is a random variable.

With respect to incident dispatching on highways, Nathanail and Zografos (1994, 1995), Zografos and Nathanail (1991) and Zografos et al (1993) have evaluated various aspects of the incident response and clearance process through analytical models, including where to locate response vehicles, which vehicles to dispatch and how to manage the process during clearance. Smith (1997) and Anderson and Fontenot (1992) examined response times and optimal vehicle positioning along linear roadways, but did not consider directional effects and interchanges, as are covered here.

Our work considers dispatching within the context of incidents that induce delays on highways. Travel time models are proposed to account for specific characteristics of highways: the side of the highway on which the incident occurs and the side of the highway on which the emergency crew is traveling; the location of interchanges at which the emergency crew can reverse direction to reach the opposite side; and the linearity of the network. In addition, we evaluate the second moment of the clearance time distribution because delay can be a quadratic function of the time required to clear an incident.

The remainder of the paper is organized as follows. First, a model is presented to illustrate the effects of randomness on incident traffic delay. Next, response time models are created for several situations in which emergency crews have different spatial distributions and in which incident induced delay may slow response to the scene. Lastly, we develop and evaluate a dispatching model that predicts average response time

as a function of the rate in which incidents occur, the service time for incidents and the number of emergency crews on duty.

The general methodology of this paper is to utilize analytical models as approximations for system performance measures. The purpose in using these models is to determine relationships between fundamental system parameters – such as the spacing between interchanges and the time to maneuver through an interchange – and system performance. In this regard, we simplify the analysis by only considering incidents that require a response from a single emergency vehicle, and do not consider multiple levels of incident priority. Future research will be needed to understand the complexities of these relationships for highway incidents. Furthermore, the actual performance of a given highway can only be determined through empirical experiments or possibly simulation. Nevertheless, the findings provide building blocks that might be used in creating more complex models.

## **2. DELAY AND THE DISTRIBUTION OF RESPONSE TIME**

Our objective is to minimize the expected traffic delay at incidents. Individual incidents are modeled in the following simplified fashion, based on constant arrival rates (see Janson and Rathi, 1991, for a model with non-constant arrival rates). When an incident occurs, queue size begins to grow at the rate  $r_g$ . When the incident is cleared, the queue size declines at the rate  $r_d$ , until such time the queue dissipates. The clearance time is the sum of four values:

- I = Incident detection time (time from incident until detection)
- W = Waiting time from incident detection until clearance vehicle is dispatched
- R = Response time from dispatch until arrival
- S = Service time to clear the incident, subsequent to arrival of response vehicle
- T = Total clearance time = I+W+R+S

The total traffic delay for an individual incident is then half the product of the maximum queue size and the total queue duration, as shown below

$$D = .5 [Tr_g][T(1 + r_g/r_d)] = .5 T^2[r_g(1 + r_g/r_d)] \quad (1)$$

For a given type of incident with known values  $r_g$  and  $r_d$ , the expected delay can be expressed as:

$$E(D) = .5 E[T^2][r_g(1 + r_g/r_d)] = .5E^2(T)[1+C^2(T)][r_g(1 + r_g/r_d)] \quad (2)$$

Where:

$$C(T) = \text{coefficient of variation for the total clearance time}$$

Equation 2 is based on the relationship  $E[T^2] = E^2(T)[1+C^2(T)]$ . It demonstrates that with constant arrival rates response crews should be dispatched in a way that minimizes the

second moment of total clearance time rather than expected total clearance time.

Alternatively, minimizing the delay can be viewed as minimizing two variables: expected total clearance time and minimizing the coefficient of variation of the total clearance time (or variance of the total clearance time). Thus queue disciplines such as “shortest-service-time-first”, which minimize expected waiting time, might not be the best for dispatching response crews as the larger variance will cause  $E(T^2)$  to increase.

### Illustration

Consider a simple case with two options for an individual incident. This example will only account for delays at the incident in question, and not for any possible effects on future incidents. Response vehicle 1 is close to the incident, but is currently busy, and response vehicle 2 is further from the incident, but is currently available. Either of the vehicles can be dispatched (vehicle 1 has a wait; vehicle 2 has no wait), but a choice cannot be revoked once it is made. Parameter subscripts denote the response vehicle, and  $S$ ,  $r_g$  and  $r_d$  are assumed to be independent of the vehicle that is dispatched. Thus:

$$\begin{aligned} E(D_1) &= \text{expected traffic delay if response vehicle 1 is dispatched} \\ &= .5 E[(I+S+W_1+R_1)^2][r_g(1 + r_g/r_d)] \end{aligned}$$

$$\begin{aligned} E(D_2) &= \text{expected traffic delay if response vehicle 2 is dispatched} \\ &= .5 E[(I+S+R_2)^2][r_g(1 + r_g/r_d)] \end{aligned}$$

We wish to find a breakeven point at which the expected delays are identical:

$$E[(I+S+W_1+R_1)^2][r_g(1+r_g/r_d)] - E[(I+S+R_2)^2][r_g(1+r_g/r_d)] = 0 \quad (3)$$

Suppose that service time is independent of the time responding to the incident, and let:

$$\begin{aligned} k_1 &= E(R_2)/E(R_1) \\ k_2 &= E(W_1)/E(R_1) \\ k_3 &= E(R_1)/E(I+S) \\ C(R_i) &= \text{coefficient of variation for } R_i \\ C(W_i) &= \text{coefficient of variation for } W_i \end{aligned}$$

Whether it is preferable to dispatch the close and busy vehicle, or the distant and available vehicle, depends on the magnitude of  $k_1$  relative to  $1+k_2$ . In the special case where  $C(R_1) = C(R_2) = 0$ , Eq. 3 can be solved as a quadratic equation, yielding:

$$k_2^* = \frac{-(1+k_3) + [(1+k_3)^2 - 4(1+k_3)(1-k_1)(k_3 C^2(W_1))]^{.5}}{2k_3 C^2(W_1)} \quad (4)$$

Eq. 4 defines a breakeven point for  $k_2$ , indicating the point where the expected delay is identical for the two vehicles. The behavior of Eq. 4 is illustrated in Figure 1, showing  $k_2^*$  as a function of  $k_1$  and  $C(W_1)$ . When  $C(W_1) = 0$ ,  $k_2^* = k_1 - 1$ , meaning that the expected response time for the more distant vehicle equals the sum of the expected response time and wait time for the closer vehicle. However, as more variation is introduced,  $k_2^*$  declines, meaning that it is better to dispatch the further vehicle, even though the expected arrival time at the scene (summing response time and waiting time) is greater. This effect is most pronounced when  $k_1$  is large, indicating that the response time for the further vehicle is large relative to the response time for the closer vehicle.

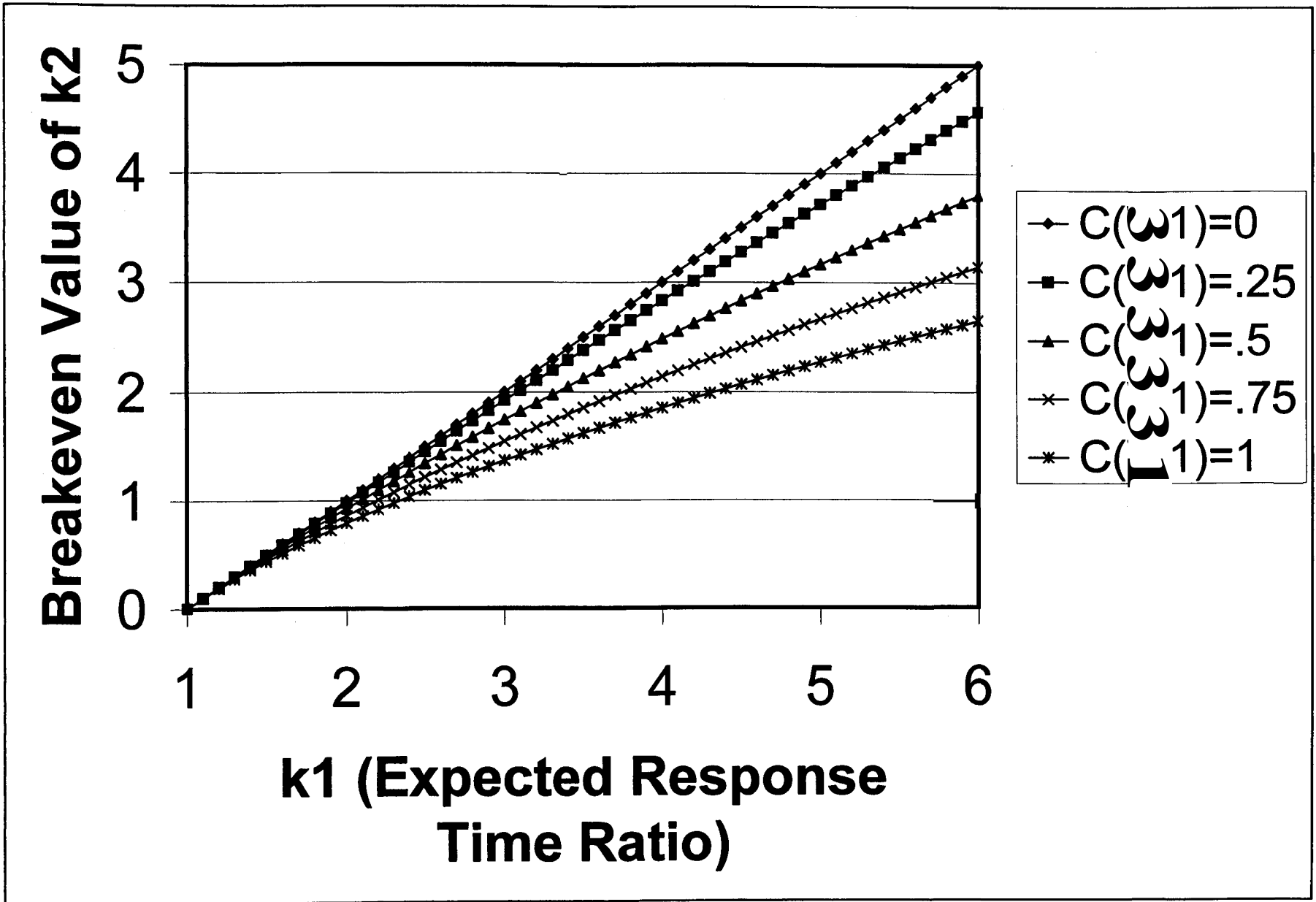


Figure 1. Breakeven value of  $k_2$  as a function of  $k_1$  and  $C(W_1)$ .



This example demonstrates that uncertainty in waiting time can make it advantageous to dispatch a more distant vehicle, even if the expected sum of waiting and response times is greater. The implications are further discussed at the end of the paper.

### **3. RESPONSE TIMES**

Incident response time is modeled here as a function the distance from the response vehicle to the incident, along with their relative direction of travel, the positioning of interchanges, and the presence of congestion that may slow incident response. We make the following assumptions: (1) The location of an incident is randomly and uniformly distributed over the length of a highway, with equal likelihood of occurring on either side; (2) Incident locations are independent of the locations of responding vehicles; (3) Each incident is served by a single vehicle; (4) Highway interchanges are spaced at a constant interval; (5) Responding vehicles can only change direction at highway interchanges. (A vehicle must be on the same side of the highway as the incident to clear the incident. Otherwise, the vehicle must change directions at an interchange to reach the incident.); (6) Speed of the response vehicle is defined by input parameters, equaling one value when the road is uncongested, and another when congested (models are provided with and without congestion).

In this section, we estimate the expectation and coefficient of variation for the response time under the following scenarios representing different dispatch strategies: (1) Each vehicle responds to incidents within its own beat, and beats are non-intersecting (i.e., each location is served by a single vehicle); (2) The closest available vehicle (in response time) is dispatched to an incident. Spacing between available vehicles is

constant and identical on both sides of the highway. The relative positions of vehicles on opposite sides of the highway are random and uniformly distributed, as vehicles are constantly in motion; (3) The closest available vehicle (in response time) is dispatched to an incident. Available vehicle locations are randomly distributed according to a stationary Poisson process.

We call strategies 2 and 3 “rolling beats”, because the territories served by vehicles are not constant. The closest vehicle (in response time) to the incident is dispatched, independent of any fixed beat assignment. Strategy 1 produces queueing, as each location can only be served by a single vehicle. Strategy 2 is advantageous from the standpoint of minimizing response time, but it is inherently unstable: as soon as a vehicle is dispatched, the separations between adjacent vehicles cannot be identical until they have time to reposition themselves. Thus, Strategy 2 is presented as a lower-bound on response time. By contrast, Strategy 3 is stable, but possibly non-optimal. By randomly transitioning a vehicle from an “available” to a “busy” state, the remaining vehicle locations will continue to be a Poisson process, with larger mean separation.

The analysis is based on the following data:

- $p'$  = time required to change directions at an interchange
- $l'$  = separation between adjacent interchanges
- $d'$  = mean distance between adjacent vehicles on each side of highway
- $v$  = velocity when traveling in uncongested traffic.

$E(R')$  = expected response time is calculated as a function of  $p', l', d'$ , and  $v'$ . To simplify results, we utilize dimensionless equations and parameters, as follows:

$$\begin{aligned}
 E(R) &= \text{expected adjusted response time} = E[R'/(d'/v')] \\
 p &= p'/(d'/v') \\
 l &= l'/d'
 \end{aligned}$$

$E(R)$  is a function of just two parameters,  $p$  and  $l$ , as the mean distance between vehicles on each side of the highway is dimensionalized to equal one, as is the velocity. Counting both sides of the highway, the mean separation between vehicles is  $\frac{1}{2}$ . Hence, under Strategy 1, each beat has a length of  $\frac{1}{2}$ . Results are first presented for the three dispatch policies assuming incident congestion does not slow response. Later, a model is presented that accounts for congestion delayed response.

### 1. Individual Beats

As shown in Figure 2, there are four ways that a vehicle can respond to an incident: (a) incident is ahead and on same side of highway; (b) incident is ahead on opposite side, (c) incident is behind and on same side, or (d) incident is behind and on opposite side. The expectations are computed as the sum of the following: linear distance traveling directly to the incident, excess distance traveling to/from an adjacent interchange, and time penalty for changing directions at the interchange. The first expectation equals  $\frac{1}{6}$  in all four cases (expectation of distance between two  $U[0, .5]$  random variables). The second expectation equals 0 in case a (interchange not used),  $l$  in

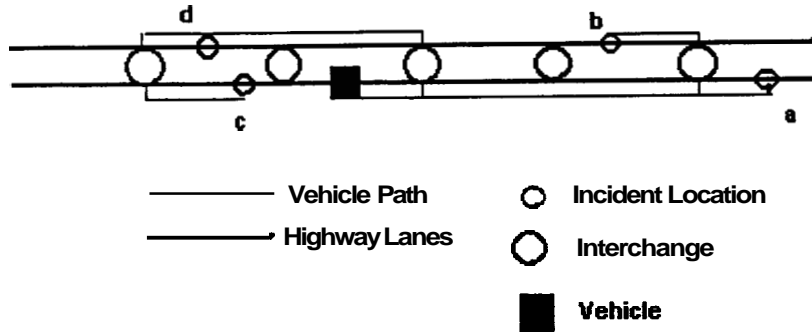


Figure 2. Vehicle can be dispatched in any of four directions depending on incident location.

cases b and c (one interchange used, with excess distance going beyond incident to interchange, then returning by equal distance), and 2l in case d (two interchanges used). The third expectation equals 0 in case a, p in cases b and c (one direction change), and 2p in case d. Hence:

$$E(R|\text{case a}) = 1/6 \quad (5a)$$

$$E(R|\text{case b}) = E(R|\text{case c}) = 1/6 + l + p \quad (5b)$$

$$E(R|\text{case d}) = 1/6 + l + l + 2p \quad (5d)$$

All cases are equally likely. Hence  $E(R) = 1/6 + l + p$ , making individual beats inefficient when interchanges are far apart or when the direction change penalty is high.

As stated earlier, the delay associated with an incident is also a function of the coefficient of variation of the clearance time. The contribution of response time depends on the second moment of the response time distribution (calculated below), as well as covariance between response time and other clearance time components (which are not evaluated here).

$$E(R^2) = .25[E(R^2|\text{case a}) + E(R^2|\text{case b}) + E(R^2|\text{case c}) + E(R^2|\text{case d})] \quad (6a)$$

$$E(R^2|\text{case a}) = E[(t_1)^2] \quad (6b)$$

$$E(R^2|\text{case b}) = E(R^2|\text{case c}) = E[(t_1 + t_2 + p)^2] = E(t_1^2) + E(t_2^2) + p^2 + 2[E(t_1 t_2) + pE(t_2) + pE(t_1)] \quad (6c)$$

$$E(R^2|\text{case d}) = E[(t_1 + t_2 + t_3 + 2p)^2] = E(t_1^2) + E(t_2^2) + E(t_3^2) + 4p^2 + 2[E(t_1 t_2) + 2pE(t_2) + 2pE(t_1) + E(t_1 t_3) + E(t_2 t_3) + 2pE(t_3)] \quad (6d)$$

Where

- $t_1$  = direct distance
- $t_2$  = excess distance at first interchange
- $t_3$  = excess distance at second interchange

The probability density function for  $t_1$  is triangular, and the probability density functions for  $t_2$  and  $t_3$  are uniform, leading to the following expectations:

$$E(t_1^2) = \int_0^5 (4-8y)y^2 dy = 1/24 \quad (7a)$$

$$E(t_2^2) = 2^2 \int_0^l y^2 / l dy = 4l^2/3 \quad (7b)$$

Under the assumption of independence stated earlier, the expressions reduce to:

$$E(R^2) = 1/24 + (11/6)l^2 + (3/2)p^2 + (2/6)l + 3lp + (2/6)p \quad (8a)$$

$$V(R) = 1/72 + (5/6)l^2 + (1/2)p^2 + lp \quad (8b)$$

Figure 3 shows that the coefficient of variation (squared) generally increases as the entrance separation ( $l$ ) increases. An exception is that when  $p=0$ , the coefficient of variation also increases when the separation declines in the range from about .1 to the vicinity of 0. The coefficient of variation is somewhat insensitive to  $p$ , though tends to decline as  $p$  increases.

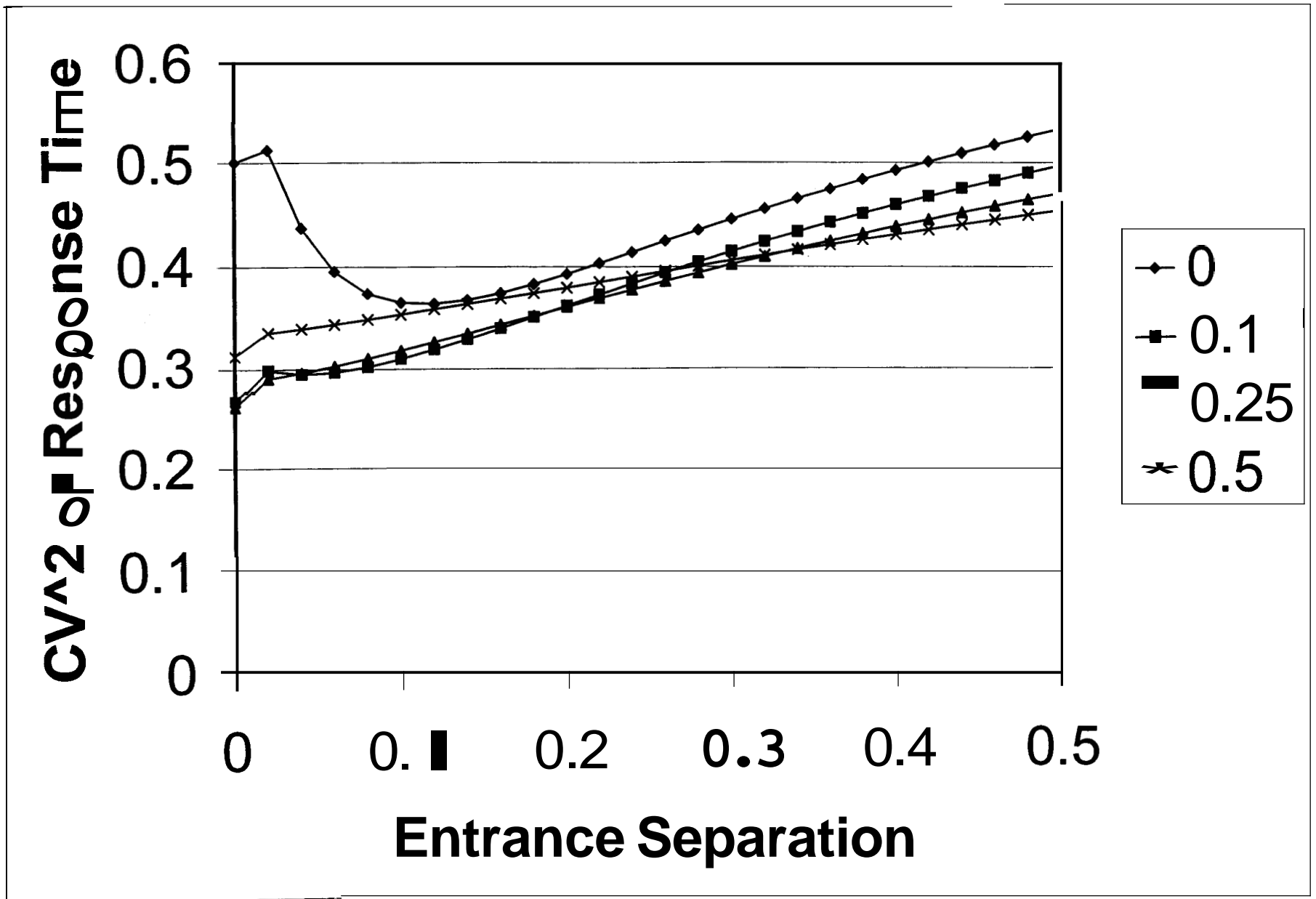


Figure 3. Coefficient of variation-squared for individual beats

## 2. Closest Vehicle/Constant Spacing

In this scenario response vehicles maintain constant spacing, though their relative position on opposite sides of the highway are constantly changing due to forward progression through traffic. Let  $y$  represent the linear distance from a vehicle to the next vehicle on the opposite side of the highway, as illustrated in Figure 4. Under rolling beats, we assume that  $y$  is a  $U[0,1]$  random variable at the time an incident is detected.

Any of four vehicles could respond to an incident: those immediately in front of, or behind, the incident, on either side of the highway. Hence:

$$R = \min\{R_1, R_2, R_3, R_4\} \quad (9)$$

Where  $R_i$  = response time for vehicle  $i$ , as numbered in Figure 4. These are calculated as follows:

$$R_1 = x_1 \quad \text{forward/same side} \quad (10a)$$

$$R_2 = x_2 + 2z_2 + p \quad \text{forward/opposite side} \quad (10b)$$

$$R_3 = (1 - x_1) + 2z_1 + 2z_2 + 2p \quad \text{backward/same side} \quad (10c)$$

$$R_4 = (1 - x_2) + 2z_3 + p \quad \text{backward/opposite side} \quad (10d)$$

Where

$x_i$  is distance from incident to vehicle  $i$

$z_1$  is distance from vehicle 3 to nearest downstream interchange

$z_2$  is distance from incident to nearest upstream interchange

$z_3$  is distance from vehicle 4 to nearest upstream interchange



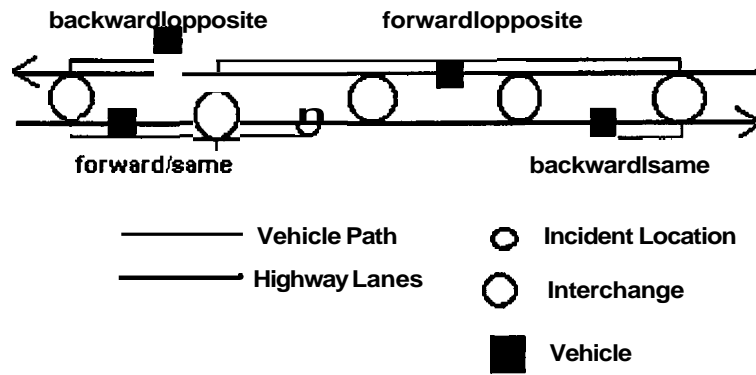


Figure 4. Incident can be served by vehicle coming from any of four directions

Expected response time is easily calculated in three extreme cases:

$$E(R) = 1/6 \quad l = p = 0 \quad (11a)$$

$$\lim_{l \rightarrow \infty} E(R) = \lim_{p \rightarrow \infty} E(R) = - \quad (11b)$$

The first case represents the expectation of the minimum of two U[0,.5] random variables, whereas the second and third cases represent the expectation of a U[0,1] random variable. Unlike individual beats, expected response time does not grow without bound as  $l$  and  $p$  increase.

The coefficients of variation are easily derived for these cases. Their squared values are shown below

$$C^2(R) = 1/2 \quad l = p = 0 \quad (12a)$$

$$\lim_{l \rightarrow \infty} C^2(R) = \lim_{p \rightarrow \infty} C^2(R) = 1/3 \quad (12b)$$

The effect of increasing the spacing between interchanges, or increasing the penalty for changing direction, is to increase the mean response time significantly. Comparing the extreme cases, it appears that the effect of increasing  $p$  and  $l$  is to decrease coefficient of variation moderately.

Simulation was used to estimate the behavior of  $E(R)$  and  $C(R)$  for more general cases of  $l$  and  $p$ , falling between the extremes (Figures 5 and 6). Each data point is based on 2000 trials, producing a standard error on the order of 1-2% of the estimated value. As either the entrance separation ( $l$ ) or penalty ( $p$ ) approaches .5,  $E(R)$  approaches the limiting value of .5. The pattern for coefficient of variation is less predictable. In general, it is insensitive to changes in parameters, but increases as either  $p$  or  $l$  increases.

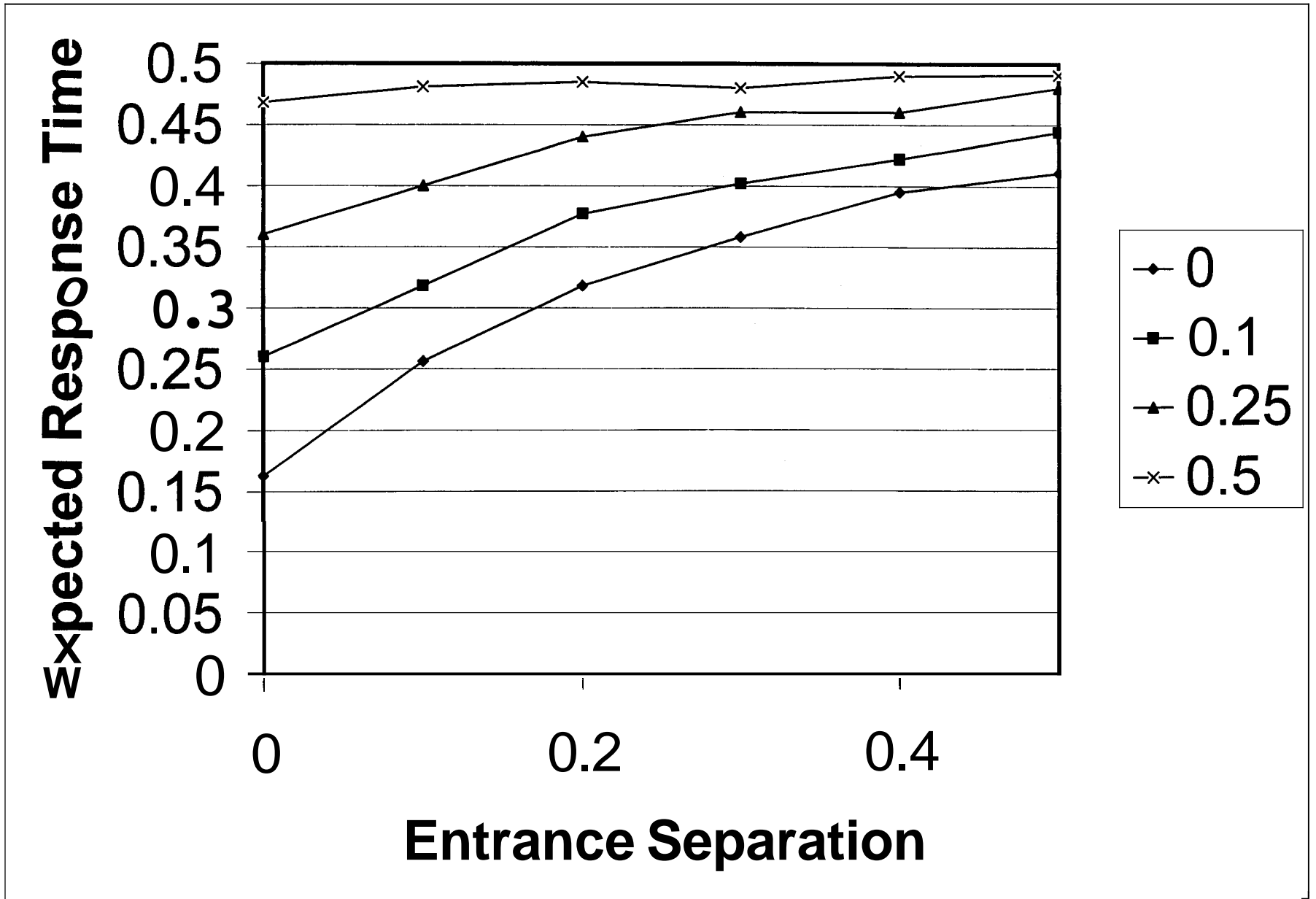


Figure 5. Expected response time for constantly spaced response vehicles as a function of entrance separation; shown for p values of 0, .1, .25, .5

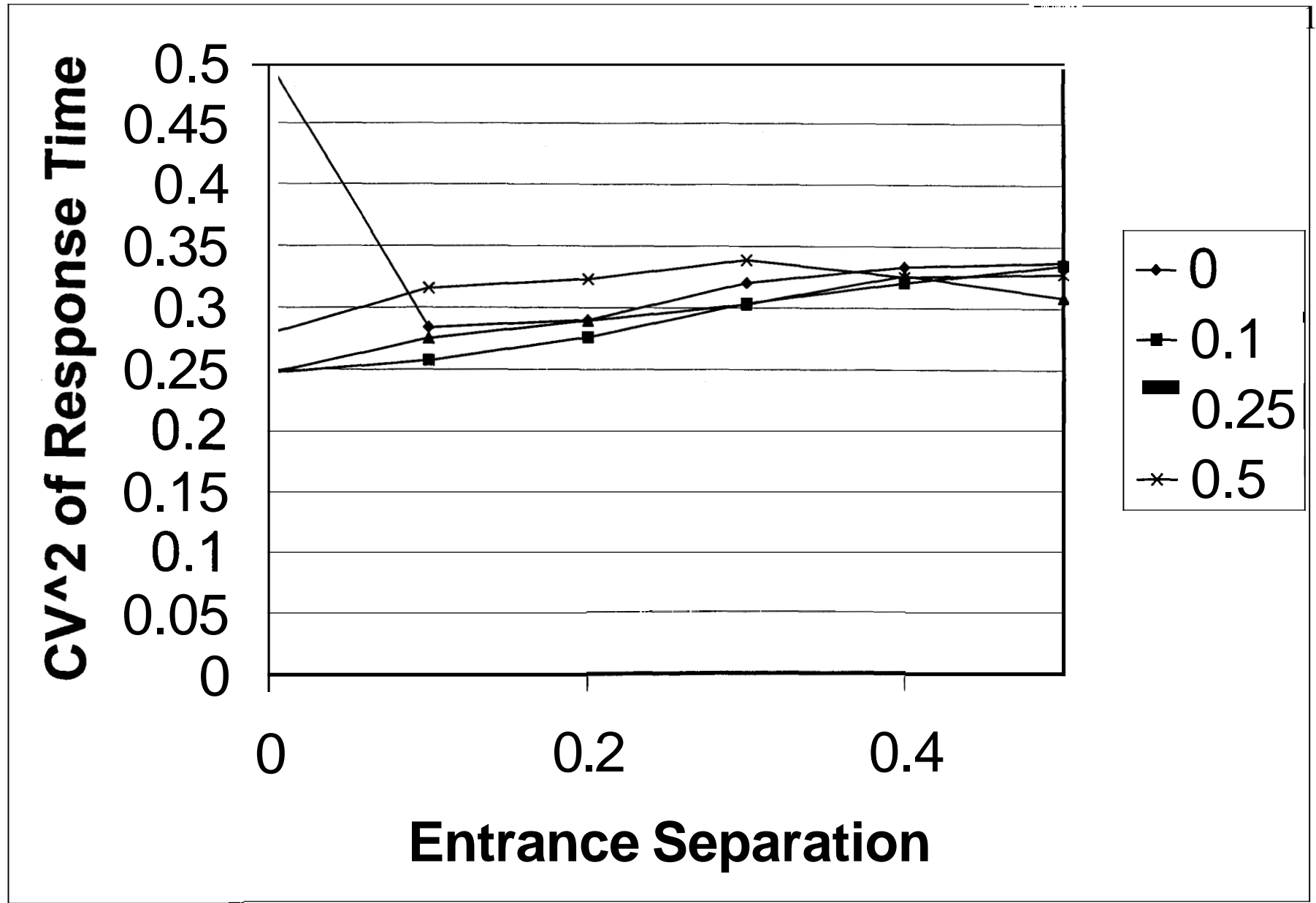


Figure 6. Squared coefficient-of-variation for response time for constantly spaced response vehicles as a function of entrance separation; shown for p values of 0, .1, .25, .5

An exception is that  $C^2(R)$  jumps to .5 when  $p=l=0$ . In this range,  $E(R)$  drops substantially, but the variance does not, resulting in an increase in coefficient of variation.

### 3. Poisson Process

As for constant spacing, the response time for Poisson process vehicle locations is defined by the minimum of four random variables, representing the response times for the adjacent vehicles on each side of the highway. Two limiting cases are easily evaluated. First, if  $l = p = 0$ ,  $E(R)$  equals the minimum of four independent exponential random variables, each representing the response time for one vehicle, and each having mean of one. The distribution for the minimum is also exponential, with  $E(R) = \frac{1}{4}$  and  $C^2(R) = 1$ . Both values are larger than the corresponding case with constant spacing.

In the other limiting case ( $p$  or  $l$  very large), vehicle 1 (same side, forward direction) always responds to the incident. Hence, the response time distribution is the distribution for a single exponential random variable, with  $E(R) = C^2(R) = 1$ . Again, the Poisson process produces larger and more variable response times. Figure 7 provides simulation results for  $E(R)$  (again with 2000 trials) for more general cases of  $l$  and  $p$ . In a qualitative sense, trends are quite similar to constant separation. However,  $E(R)$  is roughly twice as large for Poisson for all values of  $l$  and  $p$ .  $C^2(R)$  (not shown) is also roughly twice as large compared to constant separation.

### 4. Effects of Incident Induced Response Delay

In some cases upstream queueing at an incident can slow incident response. We capture this effect in the following way:

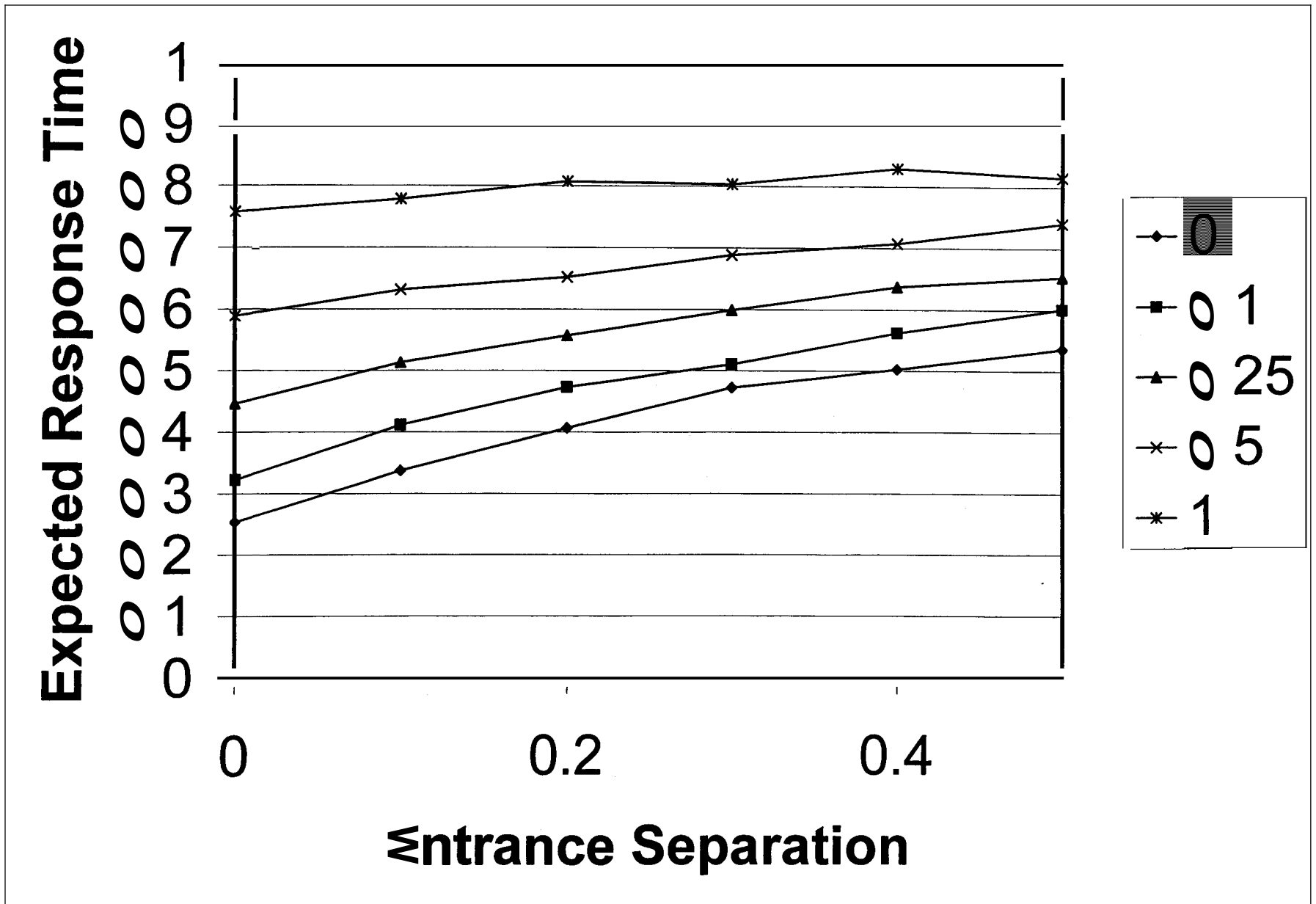


Figure 7. Expected response time for Poisson process response vehicles as a function of entrance separation; shown for p values of 0, .1, .25 and .5

- f = length of upstream queue on incident side of highway
- b = length of upstream queue on opposite side of highway
- v'' = velocity of response vehicle in queue as a ratio to v'

For example, response time for vehicle 1 (forward direction, same side), is calculated as follows:

$$R_1 = \begin{matrix} x_1/v'' & x_1 \leq f & (13a) \end{matrix}$$

$$(x_1-f) + f/v'' \quad x_1 > f \quad (13b)$$

Similar calculations were made for other response vehicles, depending on their passage through incident queues, which can potentially occur on both sides of the highway.

Simulations were completed as before. In the interest of brevity, only the second dispatch strategy was evaluated (Figures 8 and 9; rolling beats/constant separation). In the experiments, parameters were set as follows:

$$v'' = .5 \quad p = .1 \text{ or } p = .5 \quad b = 0 \text{ or } b = f/2 \quad l = .2$$

f varies from 0 to 1

With v'' = .5, an increase in f can at most double the expected response time, though the increase is usually significantly less. This is because vehicles traveling on the opposite side of the highway can circumvent some of the congestion. E(R) levels off once f = .5 for the examples. When the forward/same side vehicle is further back in the queue than the distance .5 then it is ordinarily better to dispatch another vehicle. Further increases in the queue size then have little effect on average response time. The coefficient of

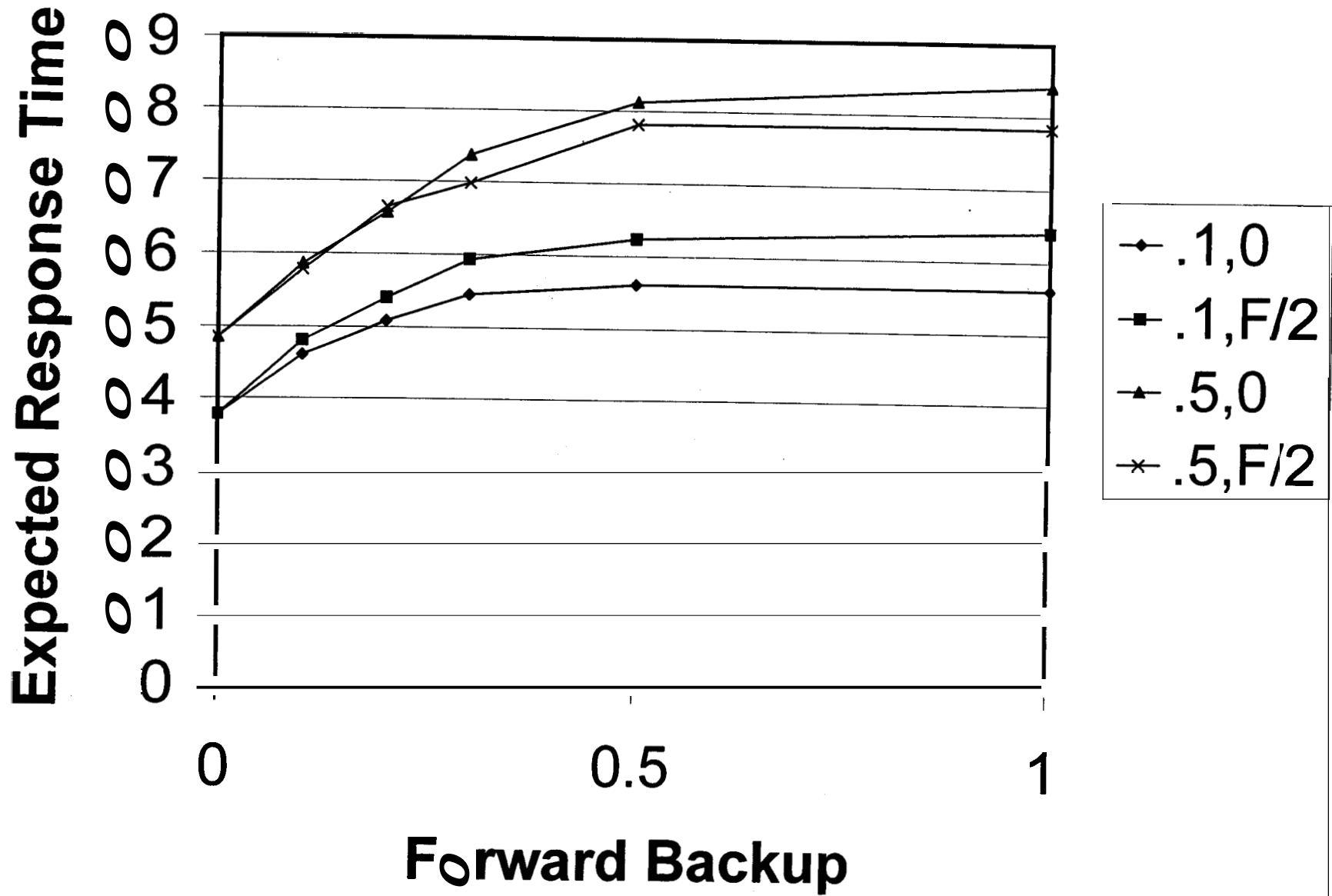


Figure 8. Expected response time with congestion delay, constant spacing;  $p = .1$  or  $.5$ ; opposite queue = 0 or 50% of same side queue



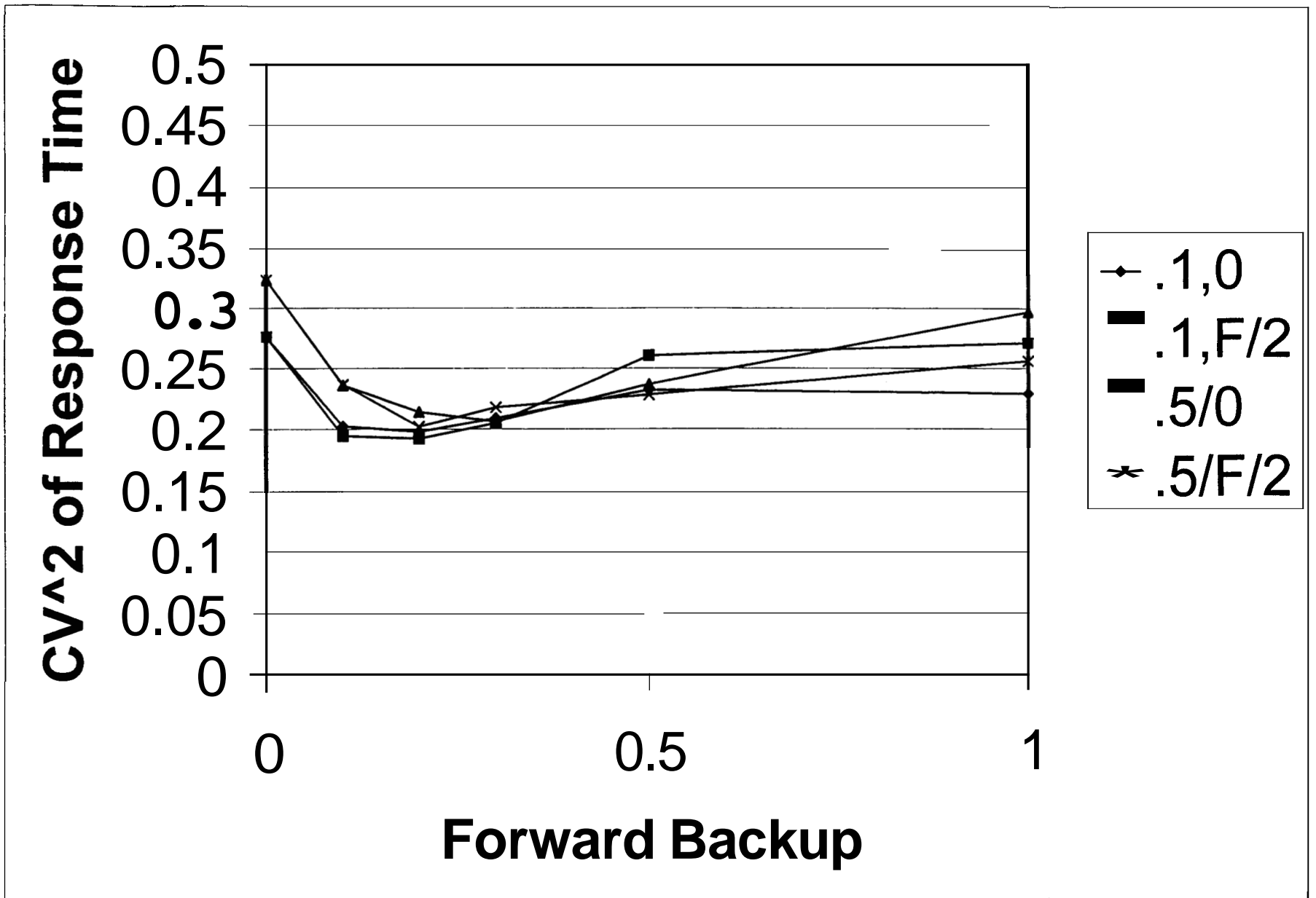


Figure 9. Squared coefficient of variation for response time with congestion delay, constant spacing;  $p = .1$  or  $.5$ ; opposite queue = 0 or 50% of same side queue

variation in response time is smallest for mid-sized backups ( $f$  approximately .25), and increases as  $f$  becomes small or large.

#### 4. EQUILIBRATION EFFECTS

The response time depends both on the number of response vehicles and the percentage of time each vehicle is busy serving incidents (time responding to the incident, i.e., travel time, plus time at the scene). In this section we estimate the expected response time accounting for the proportion of time that they are serving incidents. The method is sufficiently general to apply to all of the dispatch scenarios.

We define the busy proportion as the proportion of time vehicles are serving incidents, which is calculated as follows:

$$P = \mu[E(R) + E(S)] = \mu t \quad (14)$$

Where

$$\begin{aligned} \mu &= \text{incident arrival rate/unit distance on each side of highway} \\ E(S) &= \text{expected time serving incident at the scene, dimensionalized as a ratio to } d'/v' \\ t &= E(R) + E(S) \end{aligned}$$

$t$  is in turn approximated by:

$$t \approx k/(1-P) + E(S) \quad (15)$$

where  $k$  is the value of  $E(R)$  when  $P = 0$  (derived from Section 3). This approximation is analogous to the one used by Ignall et al (1978), who found that a square-root model approximated response distance when the number of busy vehicles randomly varies. The model is exact for locations defined by the Poisson process and approximate otherwise.

Imagine a Poisson process with rate 1 that defines vehicle locations along the side of a highway. Any individual location has a probability P of being busy. By treating this event as a Bernoulli random variable, the remaining (i.e., non-busy) vehicles are also a Poisson process, with rate (1-P).

Eqs. 14 and 15 can be solved as a quadratic equation to identify an equilibrium value of t, which we denote t\*:

$$t^* = k/(1-\mu t^*) + E(S) \quad (16a)$$

$$t^* = \frac{[1+\mu E(S)] - \sqrt{\{1-\mu E(S)\}^2 - 4\mu k}}{2\mu} \quad (16b)$$

To return a feasible solution, Eq. 16 is bounded by the following:

$$\{1-\mu E(S)\}^2 \geq 4\mu k \quad (17a)$$

$$\mu E(S) \leq 1 \quad (17b)$$

Because a square-root cannot be negative, the following upper bounds can be placed on equilibrated values of t and P:

$$t^* \leq [1+\mu E(S)]/2\mu \quad (18a)$$

$$P^* < [1+\mu E(S)]/2 \quad (18b)$$

For most values of  $\mu$  and  $E(S)$  it is impossible to attain a P value of 1, or possibly even close to one. When  $\mu E(S)$  is small, the maximum utilization is on the order of .5. When the server is busier for a greater portion of time, the system is inherently unstable: response time increases and approaches infinity as more servers become busy. These features are shown for  $t^*$  and  $P^*$  in Figures 10 and 11 for the case  $k=.4$ , with varying values of  $\mu$  and  $E(S)$ .

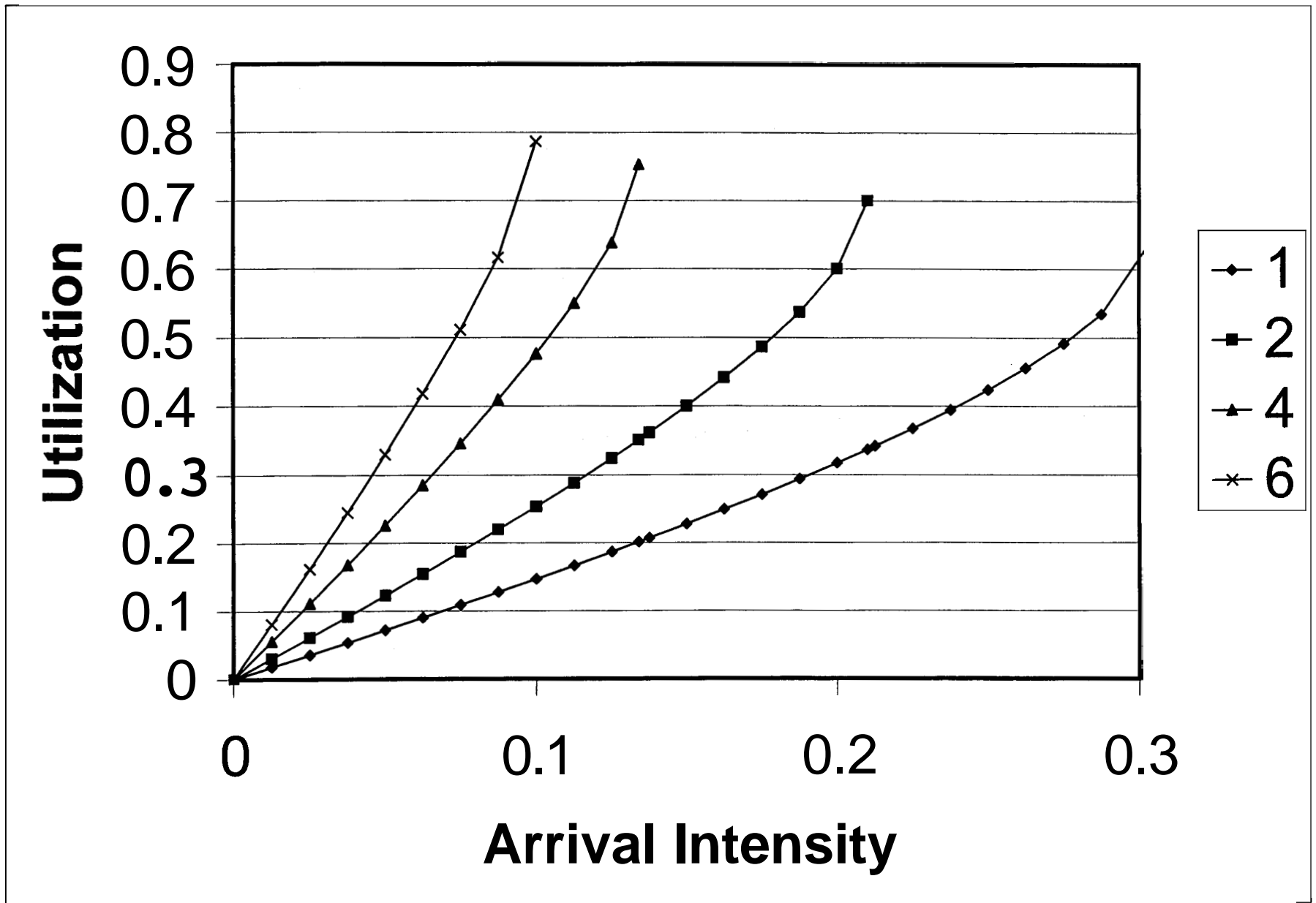


Figure 10. Equilibrated utilization as a function of arrival rate;  $k = .4$ ,  $E(S) = 1, 2, 4$  and  $6$

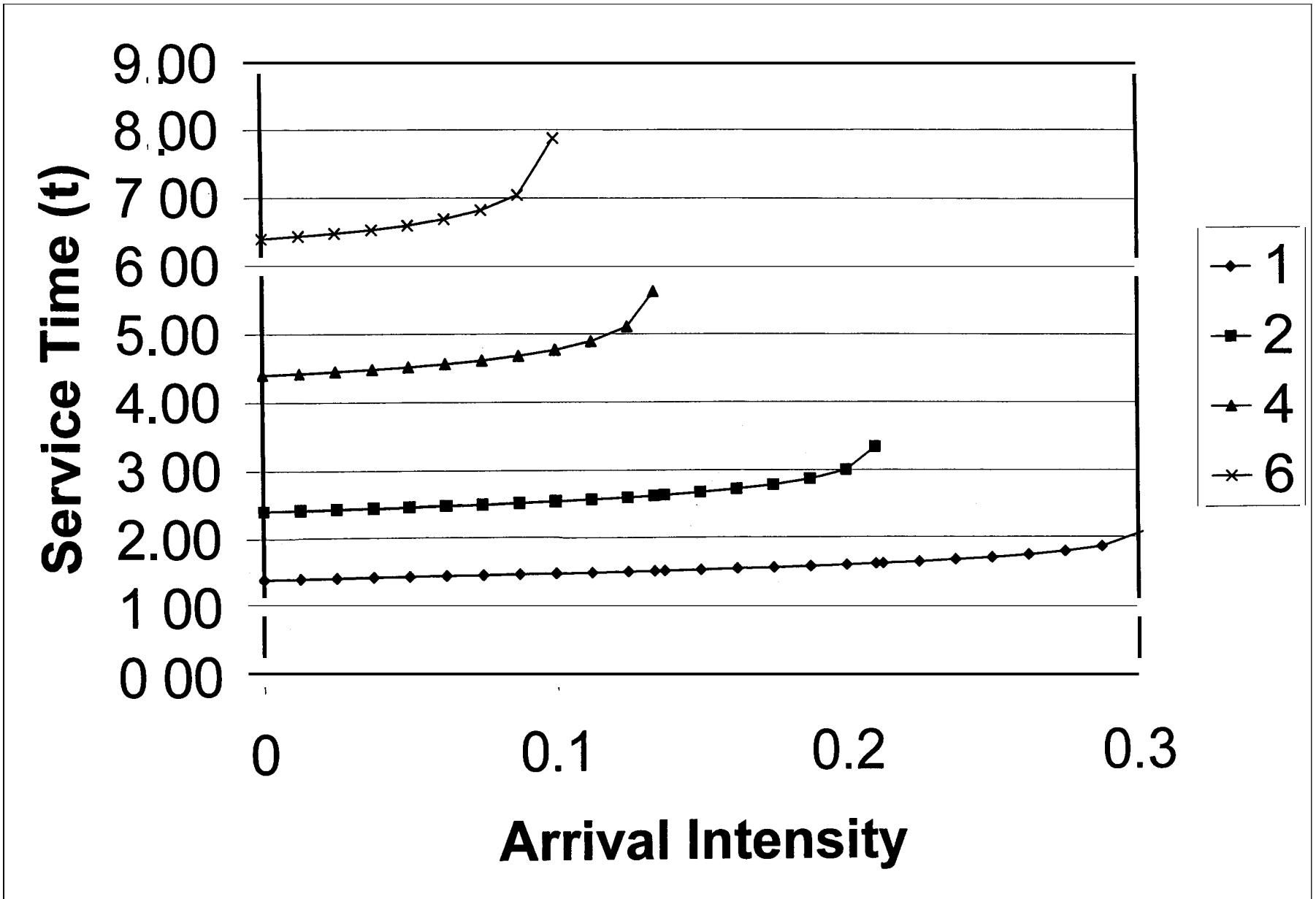


Figure 11. Equilibrated service time (t) as a function of arrival rate;  $k=4$ ;  $E(S) = 1, 2, 4$  and  $6$

From these results, Strategies 2 and 3 are only stable when the utilization is relatively low, on the order .7 or less. By contrast, if dispatch is limited to vehicles that are within a set distance of the incident, or within a vehicle's beat, response time will not grow unbounded, allowing the system to be stabilized at larger utilization levels. The drawback is that an additional delay is created, representing the waiting time from detection until dispatch (because all feasible vehicles may be busy). Nevertheless, bounding the response distance, and allowing some calls to queue, is virtually essential to keep dispatch systems stable.

## 5. SUMMARY

Returning to our initial example, the fundamental question in dispatching incident crews is whether to send the closest vehicle that is currently available or to wait for another vehicle to become available that is even closer. The advantage of waiting for a closer vehicle is that service time is effectively reduced, adding to capacity and providing stability at higher levels of utilization. On the other hand, waiting for a vehicle to become available adds uncertainty, which contributes to expected traffic delay (because traffic delay depends on the second moment of the clearance time distribution). As a consequence, any reasonably robust dispatch strategy must provide for a hybridization of the two objectives, trading-off greater certainty in response time against stability at higher levels of utilization. Individual beats provide stability at higher utilization levels, but longer waits at lower utilization levels because the assigned vehicle may be busy.

In the case of individual beats, expected response time is a linear function of both the interchange spacing and the direction-change penalty. On the other hand, with rolling

beats (either with fixed spacing or Poisson process locations), increases in interchange spacing and the direction-change penalty do not cause expected response time to increase without bound. Instead, it approaches a limit, for which the responding vehicle always reaches the incident from an upstream location on the same side of the highway. However, rolling beats, in which the closest available vehicle is dispatched to the incident, have the drawback that they become unstable more easily.

Fundamentally, any reasonably good dispatch strategy must balance the advantage of immediately dispatching a vehicle against the advantage of waiting for a closer vehicle to become available, as well as compare the travel times for vehicles on the same side of the highway to vehicles on the opposite side. This paper has taken a step in the direction of answering these questions through analysis of idealized highways.

## 6. REFERENCES

- Anderson, L.R. and Fontenot (1992). Optimal positioning of service units along a coordinate line. *Transportation Science*, 26,346-351.
- Chaiken, J.M. and Dormont, P. (1978). A patrol car allocation model: capabilities and algorithms. *Management Science*, **24**, 1291-1300.
- Green, L. (1984). A multiple dispatch queueing model of police patrol operations. *Management Science*, 30,653-664.
- Green, L. and Kolesar, P. (1984). A comparison of the multiple dispatch and M/M/c priority queueing models of police patrol. *Management Science*, 30,665-670.
- Green, L. and Kolesar, P. (1989). Testing the validity of a queueing model of police patrol. *Management Science*, 35, 127-148.

- Ignall, E.J., Kolesar, P. and Walker, W.E. (1978). Using simulation to develop and validate analytic models: some case studies. *Operations Research*, **26**, 237-253.
- Ittimakin, P.K. and Edward, P.C. (1991). Stationary waiting time distribution of a queue in which customers require a random number of servers. *Operations Research*, **39**, 633-638.
- Janson, B.N. and A. Rathi (1993). Economic feasibility of exclusive vehicle facilities. *Transportation Research Record*, N. 1305,201-214.
- Kolesar, P.J., Rider, K.L., Crabill, T.B. and Walker, W.E. (1975). A queueing-linear programming approach to scheduling police patrol cars. *Operations Research*, **23**, 1045-1062.
- Larson, R. (1972). *Police Patrol Analysis*, MIT Press, Cambridge, MA.
- Larson, R. (1974). A hypercube queueing model for facility location and redistricting in urban emergency services. *Computers and Operations Research*, **1**, 67-95.
- Larson, R.C. and Rich, T.F. (1987). Travel-time analysis of New York City police patrol cars. *Interfaces*, **17**, 15-20.
- Larson, R.C., and McKnew, M.A. (1982). Police patrol-initiated activities within a systems queueing model. *Management Science*, **28**, 759-774.
- Nathanail, T. and K. Zografos (1995). A framework for integrating real time information and decision support system for incident management operations, *International Symposium on Automotive Technology and Automation* (Stuttgart), 555-563.
- Nam, D. and Mannering, F. (2000). Hazard-based analysis of highway incident duration.



- Nathanail, T. and Kostas G. Zografos (1994). Simulation of freeway incident restoration operations. *Fifth Vehicle Navigation and Information Systems Conference* (Yokohama, Japan). 229-232.
- Smith, D.K. (1997). Police patrol policies of motorways with unequal patrol length. *Journal of the Operational Research Society*. 48,996-1000.
- Zografos, K.G. and T. Nathanail (1991). An analytical framework for minimizing freeway incident response time. *Applications of advanced technologies in transportation engineering: proceedings of the second international conference*. American Society of Civil Engineers, New York.
- Zografos, K.G., T. Nathanail, and P. G. Michalopoulos (1993). Analytical framework for minimizing freeway-incident response time, *Journal Of Transportation Engineering*. 119,535-549.