Title
Evaluation Of Highway Bottlenecks

Authors
Hall, Randolph W.
Kamoun, Mahdi

Publication Date
1991

Program on Advanced Technology for the Highway
INSTITUTE OF TRANSPORTATION STUDIES
UNIVERSITY OF CALIFORNIA AT BERKELEY

Evaluation of Highway Bottlenecks

Randolph W. Hall
Mahdi Kamoun

PATH Research Report
UCB-ITS-PRR-91-9

This work was performed as part of the Program on Advanced Technology for the Highway (PATH) of the University of California, in cooperation with the State of California, Business and Transportation Agency, Department of Transportation, and the United States Department of Transportation, Federal Highway Administration.

July 1991

ISSN 1055-1425

EVALUATION OF HIGHWAY BOTTLENECKS

Randolph W. Hall
Mahdi Kamoun
Department of Industrial Engineering and Operations Research
University of California
Berkeley, California 94720

July, 1991

SUMMARY

This is the final report of the PATH research project "Bottleneck Evaluation Model." The goal of the project was to develop a computer tool for evaluating capacity and travel time benefits of PATH improvements. In prior working papers the traffic simulation program BTS (Bottleneck *Traffic Simulator)* was documented and a preliminary analysis of new automation technologies was completed.

BTS is capable of simulating a wide range of phenomena not ordinarily included in traffic simulation programs, such as incidents and changes in traveler arrival time. With respect to incidents, BTS allows users to enter simple statistics on incident duration, magnitude and frequency. Although good data are not always available on incidents, imprecise estimates are preferable to ignoring incidents, as is customary in traffic models.

In this final report, BTS is used to investigate the time benefits of changes in highway design and operation. Key issues include the effects of (1) highway reliability, in the form of incident frequency, duration and reliability; and (2) changes in traveler behavior, in the forms of arrival time choice and reneging (travelers who skip trips due to high travel costs). Two highway segments are used as prototypes: the Caldecott Tunnel in Orinda/Oakland and the Golden State Freeway in Burbank.

Important conclusions of this report include:

- The presence of infrequent, incidents can, under some scenarios, cause large percentage increases in vehicle delay.
- Delay is highly sensitive to arrival time behavior, especially when travelers have rigid deadlines.
- Reneging has a moderating effect on delay. Failure to account for reneging will cause delay to be overestimated.
- Although incidents can be the cause of 50% or more of total delay, the effect of eliminating incidents may be similar to that of increasing lane capacity by just a few percent.

Another important conclusion is that reduction in average vehicle delay may be a misleading measure of the benefit from new highway investments. Though improvements in highway performance will surely reduce average delay, they may also enable more people to travel during peak periods. Increased traffic will undercut the reductions in delay that would occur in the absence of behavioral changes. The biggest benefit of highway investments may then be that more people have the mobility to travel during peak periods.

Independent of traffic modeling, there is a great need to improve data collection efforts before and after new highway projects. Only in this way can the benefits of highway investments be accurately assessed.

TABLE OF CONTENTS

CHAPTER 1 - BACKGROUND 1

CHAPTER 2 - HIGHWAY PERFORMANCE 3

A. Highway Incident Model 4
B. Effective Capacity 4
C. Delay Effects of Incidents 8

CHAPTER 3 - TRAVELER BEHAVIOR 12

CHAPTER 4 - EXPERIMENTAL RESULTS 14

A. Highway Incidents and Non-recurrent Congestion 15
B. Deterministic Equivalence 22
C. Changes in Arrival Pattern 26
D. Reneging 29

CHAPTER 5 - CONCLUSIONS 36

1. BACKGROUND

Congestion delays constitute one of the most significant challenges to the California highway system, especially in the urbanized areas of Los Angeles and San Francisco. Delays are costly in terms of driver time, vehicle operating expense and air pollution. Delays also add to the cost of doing business in California, both in terms of logistics and in terms of higher wages paid to employees in compensation for long commutes.

Traditionally the problem of congestion has been addressed by adding lanes to the highway system or through expanding mass transit, but the advent of advanced highway technologies has opened new possibilities. Systems for highway automation may enable highway capacity to be expanded without adding lanes. Highway performance may also be improved through incident management strategies that reduce the duration, frequency and magnitude of highway incidents, or congestion management schemes that automatically control traffic signals in response to changing conditions.

Technology might also be used to induce changes in traveler behavior. Potentially, small reductions in vehicular traffic over congested roads can translate into large improvements in delay. Changes in travel frequency, time, route and mode can all work toward this end. Changes in trip origin (e.g., moving closer to work) and destination (e.g., changing to a job closer to home) can also be beneficial. In the future, new technologies for automatic toll collection and roadway pricing, and traveler information, might effect these improvements .

The objective of the project “Bottleneck Evaluation Model” has been to develop methodologies for assessing the benefits of new transportation technologies. To this end, prior working papers accomplished the following:

"Bottleneck Traffic Simulator (BTS) Version 1.1" documented a macroscopic computer program for simulating delays at highway bottlenecks. BTS simulates the random occurrence of traffic incidents and weather patterns, calculates mean vehicle delay, and creates graphs of delay as a function of time. BTS also incorporates a behavioral component to simulate traveler response to highway delays in the forms of: (1) when to travel, (2) route to travel, and (3) whether. to travel (Lin and Hall, 1991).

"Time Benefits of New Transportation Technologies" presents new concepts for highway automation as well as methodologies for assessing time benefits. These automation concepts include small-scale projects where streams of vehicle traffic are automatically merged at slow speeds. Another concept is the "mini-highway", whereby vehicles operate at slow speeds, but high capacity, within narrow urban corridors (Hall, 1991).

In this final report, BTS results are presented for a prototype highway bottleneck. These simulations both illustrate the capabilities of BTS and provide insights into the effects of changes in highway performance. The simulations also point to opportunities for new research, as will be discussed in the conclusions .

The remainder of the paper is divided into four chapters. The first develops analytical models of highway performance as a function of incident frequency, magnitude and duration. The second presents models of traveler behavior. Next, the results from a series of simulation experiments are provided. The final chapter offers conclusions.

CHAPTER 2. HIGHWAY PERFORMANCE

New technologies can improve the performance of highways through changes in highway design (e.g. , automation) , highway operation (e.g. , congest ion management) or traveler behavior (e.g., traveler information). Though the benefits of new technologies cover the spectrum from enhanced driver comfort to added safety, the focus of this paper is on travel time improvements.

In this report, the time required to travel between two points on a highway is viewed as the sum of two factors: a free-flow time (travel time in the absence of congestion) and a congestion delay. Free- flow time is primarily a function of design geometrics and associated speed limits, as well as trip circuity. Congestion delay is the sum of a recurrent component and a non- recurrent component, the former being a function of normal highway capacity and traffic patterns, and the latter being dependent on incidents and other random events . Combining all three elements:

$$T = T_f + [T_{cr} + T_{cn}] , \tag{1}$$

where: T_f = free-f low travel time

T_{cr} = recurrent congestion delay

T_{cn} = non-recurrent congestion delay.

Improvements in the travel time performance can occur in any of the three elements, for instance: T_f could be reduced through construction of faster roads or shorter routes; T_{cr} could be reduced through instituting tolls that discourage peak period travel; and T_{cn} could be reduced through improved highway safety that reduces the number of incidents.

BTS models delays at highway bottlenecks through use of a queueing model with random capacity. The delay equals the time spent in queue at the entrance to the bottleneck. Average recurrent delay is the average time spent in queue in the absence of incidents, and average non-recurrent delay is the average additional time in queue due to incidents and other random events.

A. Highway Incident Model

Within BTS, congestion delay is computed as a function of the highway capacity and the vehicle traffic pattern. The capacity can either be the normal capacity or a reduced capacity if incidents are present. Incidents are viewed as random events that occur with a set probability per unit time. Between the time an incident occurs and the time it is cleared, there may be an increased likelihood of further incidents as drivers encounter hazardous driving conditions. In the presence of multiple incidents, the reduced capacity is defined by the most significant incident; that is, the outstanding incident that causes capacity to be reduced by the largest percentage.

B. Effective Capacity

Effective capacity will refer to the average roadway capacity after accounting for the random occurrence of incidents. Let :

p_1 = probability of a new incident in a time slice, given no current incident.

p_2 = probability of a new incident in a time slice, given a current incident.

C = roadway capacity in the absence of incidents

$\bar{M}$ = average incident magnitude (capacity loss as a proportion)

$\bar{D}$ = average incident duration (in minutes)

w = length of a time slice (in minutes).

Because multiple incidents can be present in the same time slice, it is difficult to develop an exact formula for effective capacity. On the other hand, by introducing reasonable approximations, simple upper and lower bounds can be created.

A lower bound is found by assuming that every incident has the effect of reducing capacity by $(C\bar{M})\bar{D}$. Consider the special case where $p_1 = p_2$. Then:

$$E(C) \geq \tilde{C} \cdot [1 - \bar{M}\bar{D}(p_1/w)] \quad . \qquad (2)$$

In reality, the mean effect of an incident is somewhat less than $\tilde{C}\bar{M}\bar{D}$. In the presence of simultaneous incidents (i.e., when a second incident occurs before the first is cleared), the loss in capacity is not the sum of the incident magnitudes but their maximum. However, so long as incidents tend not to be present simultaneously, the approximation is accurate.

Eq. 2 can be generalized to include the case $p_1 \neq p_2$. Let p' represent the average probability of a new incident in a time slice, accounting for periods when there are incidents and periods when there are no incidents. Let:

P_1 = probability of no outstanding incidents in a time slice

$$\geq 1 - \frac{p'\bar{D}}{w} \tag{3a}$$

P_2 = probability of an outstanding incident in a time slice

$$\leq \frac{p'\bar{D}}{w} . \tag{3b}$$

Eqs. 3a and 3b are inequalities because they do not account for simultaneous incidents.

p' can be derived from P_1 and P_2. If we assume that $p_2 \geq p_1$ (meaning driving is more hazardous in the presence of incidents), the following bounds are obtained:

$$p' = p_1 P_1 + p_2 P_2 \tag{4a}$$

$$\leq p_1 (1 - p'\bar{D}/w) + p_2 (p'\bar{D}/w) \tag{4b}$$

$$\leq \frac{p_1}{1 - (p_2 - p_1)(\bar{D}/w)} . \tag{4c}$$

The bound on expected capacity becomes:

$$E(C) \geq \tilde{C} \cdot [1 - \overline{MD}(p'/w)] . \tag{5}$$

An upper bound on capacity is found by assuming that incidents exhibit a "blocking mechanism." That is, a second incident cannot occur until the first is cleared (i.e., $p_2 = 0$). Let:

N = expected length of a "normal" period
(interval without incidents)

I = expected length of an "incident" period
(interval with incident, or incidents, present).

The length of a normal period is a geometric random variable, whose expectation is given by:

$$N = \frac{w(1-p_1)}{p_1} . \tag{6}$$

The probabilistic structure of an incident period is more complicated. However, by assuming that $p_2=0$, the following lower bound is obtained:

$$I \geq \bar{D} . \tag{7}$$

The effective capacity is now bounded by the following:

$$E(C) \leq \tilde{C}\left[1 - \bar{M}\frac{I}{I+N}\right] = \tilde{C}\left[1 - \bar{M}\bar{D}\left[\frac{p_1}{p_1\bar{D}+(1-p_1)w}\right]\right] . \tag{8}$$

The primary difference between Eq. 8 and Eq. 5 is that the term $-p_1 w$ appears instead of $-p_2\bar{D}$ in the denominator of the second term, which yields a somewhat larger value for E(C).

A second- order upper bound on E(C) was also created by ad just ing I to account for the possibility of a second incident before the first is cleared, but not a third. In such an event, the incident period is extended and the expected capacity is reduced accordingly. Though the result is used to run simulations later in the paper, it will not be presented here because it cannot be expressed in closed-form.

C. Delay Effects of Incidents

When an incident occurs, vehicles incur non- recurrent congest ion delays. This added. delay comprises an immediate (during the incident) delay and a residual (after the incident is cleared) delay. The following variables are used in modeling these delays:

$D(t)$ = cumulative vehicles to depart from bottleneck by time t in the absence of incident

$D_i(t)$ = cumulative vehicles to depart in the presence of an incident

a = time that incident begins

e' = first time that queue vanishes after incident is cleared.

Then:

$$W = \text{extra delay due to incident} = \int_a^{e'} [D(t)-D_i(t)]\,dt \, . \tag{9}$$

An incident can either occur at a time when a queue of vehicles already exists or at a time when a queue does not exist. Consider the first case. Suppose that the incident has a duration d, a magnitude m, and that the incident is cleared before the queue would ordinarily vanish, denoted time e. Referring to Figure 1, during the period from a to $a+d$, $D(t)-D_i(t)$ grows at

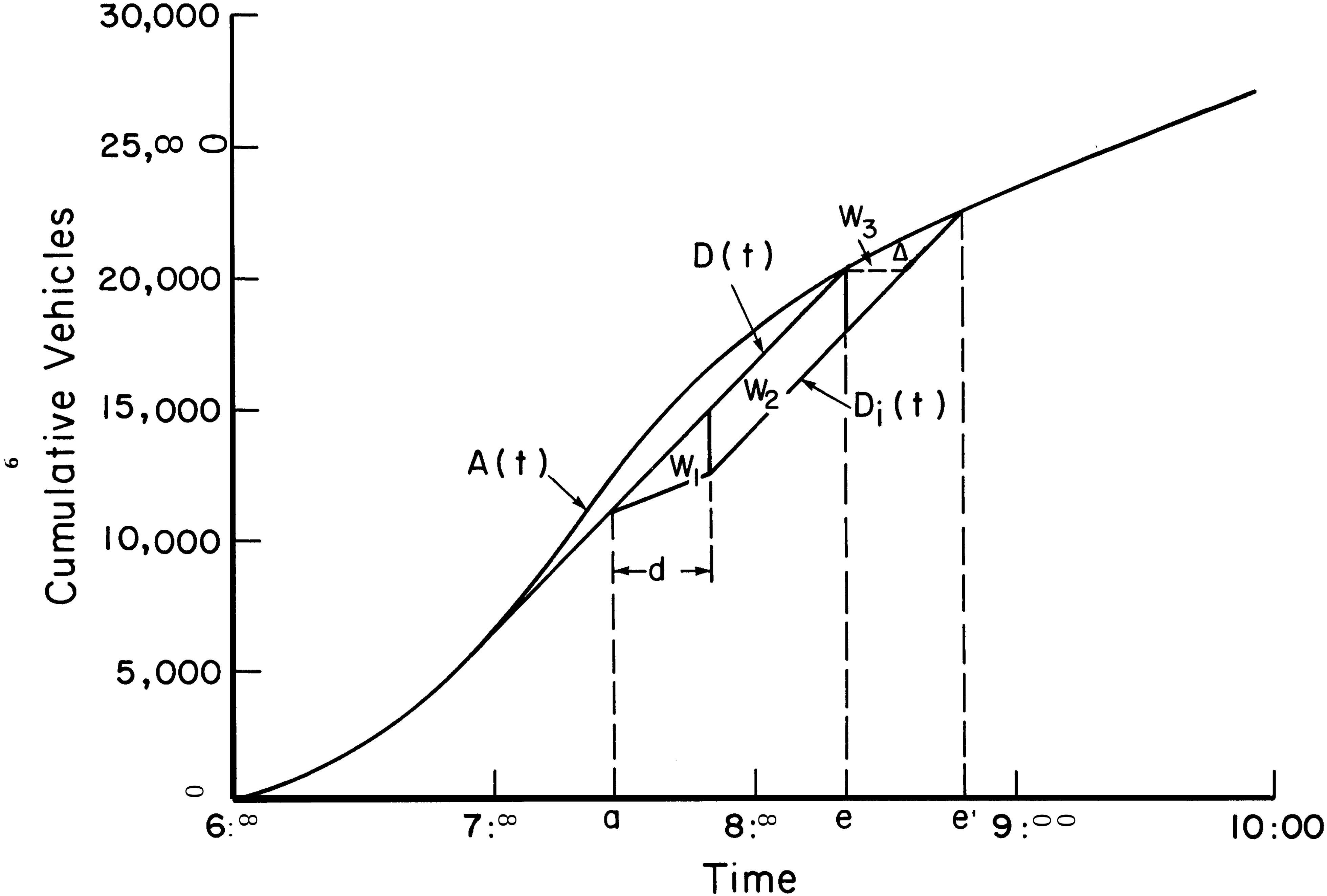

Figure . Added delay from an incident includes an immediate component (W_1) and residual components (W_2 + W_3).

the rate Cm and reaches a peak of Cmd at time a+d. At this point, the incident is cleared and capacity returns to normal. Between time a+d and e, $D(t)-D_i(t)$ remains at Cmd in the form of a residual queue. Finally, between time e and e' the residual queue dissipates at the rate $\tilde{C}-\lambda(t)$, where $\lambda(t)$ is the arrival rate at time t. Combined, the added waiting time amounts to:

$$W = d(\tilde{C}md/2) + (e-a-d)(\&d) + \int_e^{e'} [A(t)-D_i(t)]dt \tag{10}$$
$$= W_1 + W_2 + W_3 .$$

$\lambda(t)$ can never be less than zero. Therefore, W_3 is bounded by the area of the triangle shown in Fig. 1, and:

$$W_3 = (\tilde{C}m^2 d^2)/2 + A , \tag{11}$$

where A is the positive difference between W_3 and the area of the triangle. Combining all terms:

$$W = [\tilde{C}md] [(e-a) + (d/2) (m- 1)] + A . \tag{12}$$

Eq. 12 indicates that the approximate effect of an incident is to increase the queue size by Cmd over the period from time a to e. From the formula, it is clear that an incident has the largest effect if it occurs at the start of the rush- hour. Nevertheless, Eq. 12 is only a rough approximation because A is highly sensitive to the arrival pattern after time e. If traffic volumes only decline gradually at the end of a rush hour, then the residual effects can persist long after time e, and A can be quite large. On the other hand, if traffic volumes exhibit a sharp peak, then the residual effects, and A, will be small.

The effect of an incident is much smaller if it occurs when a queue would not otherwise exist, especially if the highway capacity is much greater than the arrival rate. If the magnitude of the incident is less than the capacity surplus, no queue would form. But even if a queue materializes, it will begin to dissipate as soon as the incident is cleared (provided there is still surplus capacity), unlike the case of Figure 1. As a consequence, it is much more important to keep a highway free of incidents during rush-hours than during the off- peak.

In general, the presence of incidents has a somewhat larger effect on waiting time than an equivalent deterministic change in capacity. (This is a consequence of Jensen's inequality, and the fact that the relationship between delay and capacity is convex.) However, it is difficult to precisely characterize the effects of incidents with simple formulas, or with simple functions of the effective capacity, especially when travelers change their behavior in response to incidents. This is the motivation for BTS, which estimates the non- recurrent component of congest ion delay through random simulation of incidents .

CHAPTER 3. TRAVELER BEHAVIOR

It is unrealistic to assume that traffic patterns will remain static in the face of changes in highway performance. If delay increases, drivers may choose to avoid peak periods, select another route, or skip the trip altogether. If capacity is expanded and delay is reduced, drivers may return to traveling in the peak period. As a consequence, traffic levels may increase and the net travel time reduction will be less than expected.

BTS provides a new approach for modeling traveler behavior based on a dynamic representation of traffic patterns. Unlike traditional approaches, BTS does not seek to find a traffic equilibrium. Instead, BTS models the evolution of travel behavior over time as the base traffic volume changes. BTS assumes that drivers annually survey the traffic conditions experienced over the preceding year. Based on average travel times, or percentile points of the travel time distribution, drivers reoptimize their choices, to be executed in the following year. Finally, BTS increments the base traffic volume in the following year according to historical growth patterns.

BTS allows travelers to respond to delays by changing their arrival times or routes, or by "reneging." A renege occurs when a traveler chooses not to travel at all because the cost of travel (in terms of delay costs and, perhaps, tolls) has become too large. BTS allows for three types of traveler behavior:

Type 1:
  Arrival Time: Fixed and invariant to delays.
  Route: Shortest travel time.
  Reneging: Not allowed

Type 2:

Arrival Time: Fixed time by which they need to depart from the bottleneck. Travelers choose to arrive at the latest possible time that assures they will depart on-time with a set probability.

Route: Travelers select the route that allows them to depart at the latest time while meeting the probability criterion.

Reneging : Not allowed.

Type 3:

Arrival Time: Travelers have a preferred time to depart from the bottleneck. However, they arrive at the time that minimizes their personal cost. The personal cost includes a travel time cost, an earliness time cost (if departure time is earlier than desired) and a lateness cost (if departure time is later than desired), plus any road toll, if applicable.

Route : Travelers select the route that minimizes personal cost.

Reneging : A percentage of travelers reneges when the minimum travel cost exceeds threshold values.

Although BTS is capable of analyzing all of the above phenomena, the experiments conducted in Chapter 4 only account for Type 1 and Type 3 behavior and a single route.

Over time, changes in traveler behavior can either accentuate or reduce delays. For instance, if travelers face rigid deadlines, then the peaks in arriving traffic volume may be magnified as travelers compete to arrive at the best possible times. On the other hand, if travel delays cause the cost of travel to increase, then fewer people will travel. The effect will be to hold delays down.

To investigate these issues in detail, the following chapter provides simulation results covering a range of traveler behavior for highways with different levels of reliability.

CHAPTER 4. EXPERIMENTAL RESULTS

BTS was used in a series of experiments to estimate the effects of changes in bottleneck design and operation. Experimental data was drawn from two highway segments: the Golden State Freeway (I- 5) northbound in Burbank, California, and the Caldecott tunnel (24) westbound in Orinda, California. The former was selected because it does not currently experience significant congest ion, so traveler arrival patterns reflect free- flow conditions. The latter is a prominent highway bottleneck. In both cases, experiments covered the morning commute , from 6:00 a.m. to 9:30 a.m.

Highway capacities approximated current conditions, 8,800 vehicles/hour in the case of Caldecott and 8,000 vehicles/hour in the case of I-5. Data was unavailable on the frequency, magnitude and duration of incidents. Therefore, sensitivity analyses were completed covering a wide range of conditions, as defined by the following:

- Incident duration has a discrete uniform distribution over [1,b], as measured in time slices of five minutes.
- All incidents have the same magnitude, expressed by the proportion M.
- Incident probabilities are defined by p_1 and p_2.

In each case, BTS was run for 10 iterations, with an increase in base traffic volume of 5% per iteration. Because Caldecott already experiences significant congest ion, the base traffic volume was reduced by 22% for the first iteration. This meant that the sixth iteration of BTS corresponded to current conditions. For both sites, it took about six iterations of BTS before recurrent congestion appeared. Cumulative base arrival curves are shown for the two sites in Figures 2 and 3 for Iterations 2, 4, 6, 8 and 10.

A. Highway Incidents and Non-recurrent Congestion

In the first set of experiments, all travelers were Type 1 (arrival pattern independent of delays), but incident parameters b, M, and p_1 were varied. p_2 was initially set equal to p_1:

1a) p_1=.01, M=.2, b=3	1b) p_1=.01, M=.2, b=12
2a) p_1=.03, M=.2, b=3	2b) p_1=.03, M=.2, b=12
3a) p_1=.01, M=.5, b=3	3b) p_1=.01, M=.5, b=12
4a) p_1=.03, M=.5, b=3	4b) p_1=.03, M=.5, b=12 .

Results (average and 95th percentile of sum of recurrent and non-recurrent delay) are provided for I- 5 in Table 1 and for the Caldecott Tunnel in Table 2. That congestion can increase rapidly as traffic grows is illustrated by the progression in delay between iterations. For instance, delay in the 10th iteration varied from 60 to almost 400% larger than the 8th, even though traffic increases by little more than 10%. The example also illustrates that small changes in effective capacity can bring about large changes in delay when operating near or above capacity. In the most extreme case of Table 1, Example 4b, effective capacity is 7250/hour, yet relative to Example la, with a capacity of 7970/hour (10% greater), delay is a minimum of 140% larger.

The rapid growth in delay is further illustrated in both Figures 4 and 5 (Examples lb and 2b). As the total traffic volume increases, both the

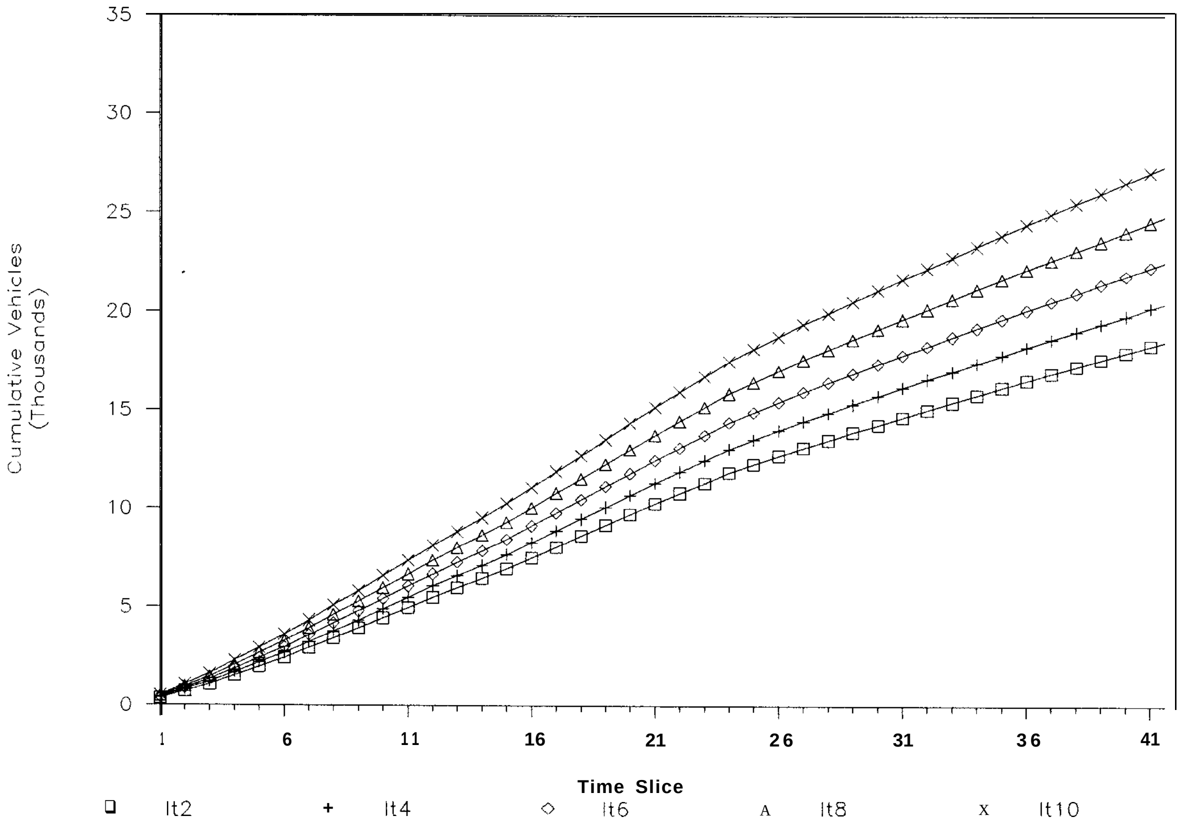

Figure 2. Cumulative arrivals at Interstate 5, Burbank **(6/12/90),** with 5% growth between iterations.

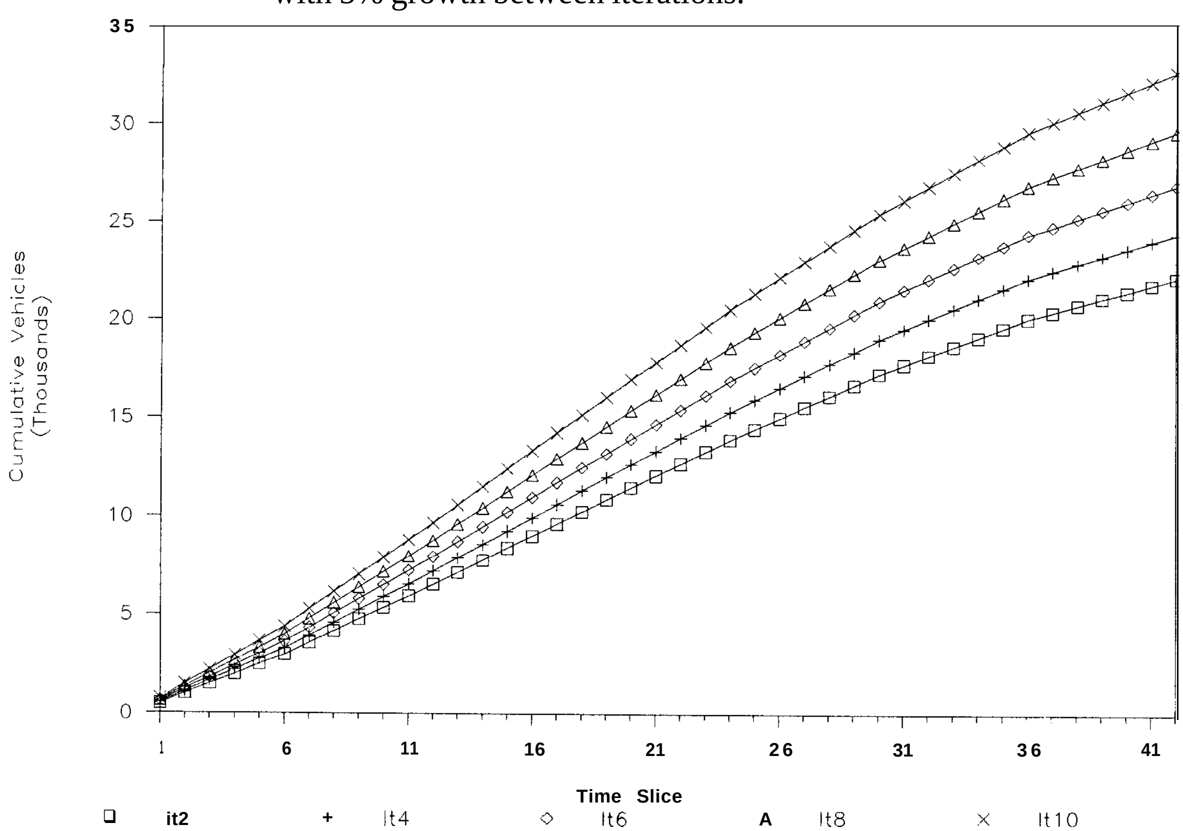

Figure 3. Cumulative arrivals at Route 24, Caldecott Tunnel It. 6 is **4/23/87,** with 5% growth between iterations.

Table 1. Time in Queue at I-5 (Minutes)
Type 1 Travelers / p1 = p2

Example 1 : *M = 0.2 , p1 = p2 = 0.01*

	b = 3		b = 12	
	Avg Q	95%	Avg Q	95%
Iter 2	0	0	0.01	0.05
Iter 4	0	0.01	0.04	0.16
Iter 6	0.02	0.23	0.26	1.83
Iter 8	1.24	2.41	1.8	6.22
Iter 10	6.79	8.74	7.74	13.82

Example 2 : *M = 0.2 , p1 = p2 = 0.03*

	b = 3		b = 12	
	Avg Q	95%	Avg Q	95%
Iter 2	0	0.01	0.02	0.17
Iter 4	0.03	0.23	0.23	1.29
Iter 6	0.14	0.81	0.75	4.03
Iter 8	1.65	3.55	3.32	9.86
Iter 10	7.44	10.51	9.59	18.63

Example 3 : *M = 0.5 , p1 = p2 = 0.01*

	b = 3		b = 12	
	Avg Q	95%	Avg Q	95%
Iter 2	0.06	0.05	0.57	4.32
Iter 4	0.09	0.27	0.78	6.28
Iter 6	0.28	1.87	1.62	11.27
Iter 8	1.75	5.45	3.84	18.73
Iter 10	7.6	12.7	10.25	27.16

Example 4 : *M = 0.5 , p1 = p2 = 0.03*

	b = 3		b = 12	
	Avg Q	95%	Avg Q	95%
Iter 2	0.16	1.36	1.81	11.3
Iter 4	0.42	2.78	3.55	16.51
Iter 6	0.75	4.48	4.75	22.41
Iter 8	3.22	9.72	9.71	32.99
Iter 10	9.34	17.62	15.55	41.76

Table 2. Time in Queue at Caldecott (Minutes)
Type 1 Travelers / p1 = p2

Example 1 : *M = 0.2, p1 = p2 = 0.01*

	b = 3 Avg Q	b = 3 95%	b = 12 Avg Q	b = 12 95%
Iter 2	0	0	0.02	0.17
Iter 4	0.01	0.03	0.13	0.98
Iter 6	0.38	1.33	0.82	4.48
Iter 8	5.52	7.32	6.36	12.43
Iter 10	13.62	15.7	14.63	21.2

Example 2 : *M = 0.2, p1 = p2 = 0.03*

	b = 3 Avg Q	b = 3 95%	b = 12 Avg Q	b = 12 95%
Iter 2	0.01	0.03	0.06	0.51
Iter 4	0.07	0.51	0.6	3.02
Iter 6	0.63	2.29	1.92	7.82
Iter 8	6.26	9.01	8.68	16.77
Iter 10	14.37	17.67	16.66	26.32

Example 3 : *M = 0.5, p1 = p2 = 0.01*

	b = 3 Avg Q	b = 3 95%	b = 12 Avg Q	b = 12 95%
Iter 2	0.09	0.17	0.81	6.39
Iter 4	0.15	0.78	1.21	9.4
Iter 6	0.78	3.8	2.59	15.41
Iter 8	6.26	11.3	8.67	25.3
Iter 10	14.47	19.79	17.27	34.98

Example 4 : *M = 0.5, p1 = p2 = 0.03*

	b = 3 Avg Q	b = 3 95%	b = 12 Avg Q	b = 12 95%
Iter 2	0.25	2.22	2.67	15.65
Iter 4	0.69	3.92	5.16	21.76
Iter 6	1.75	7.38	7.04	28.55
Iter 8	8.51	16.43	15.82	40.97
Iter 10	6.4	24.94	22.94	49.85

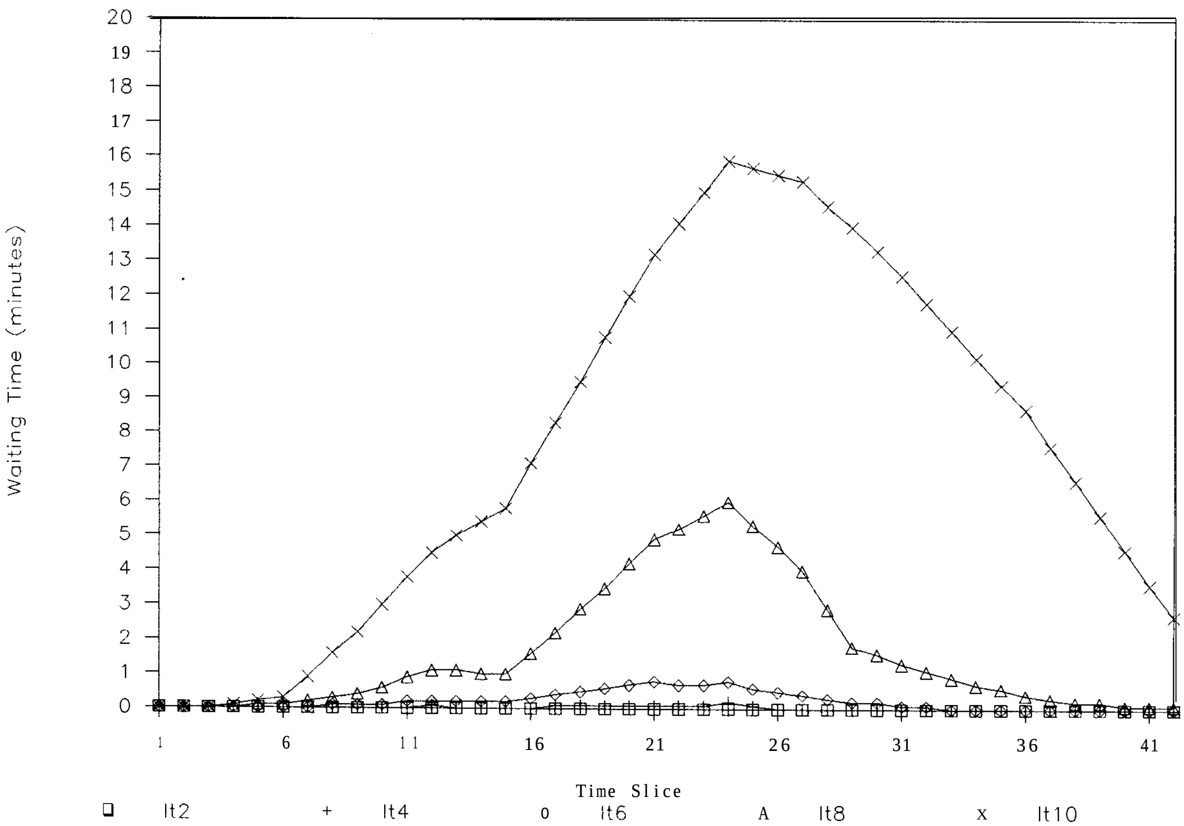

Figure 4. **Average time in queue versus arrival time at I-5 with Type 1 travelers, Example lb (M=.2, b=12, p_1=p_2=.01)**

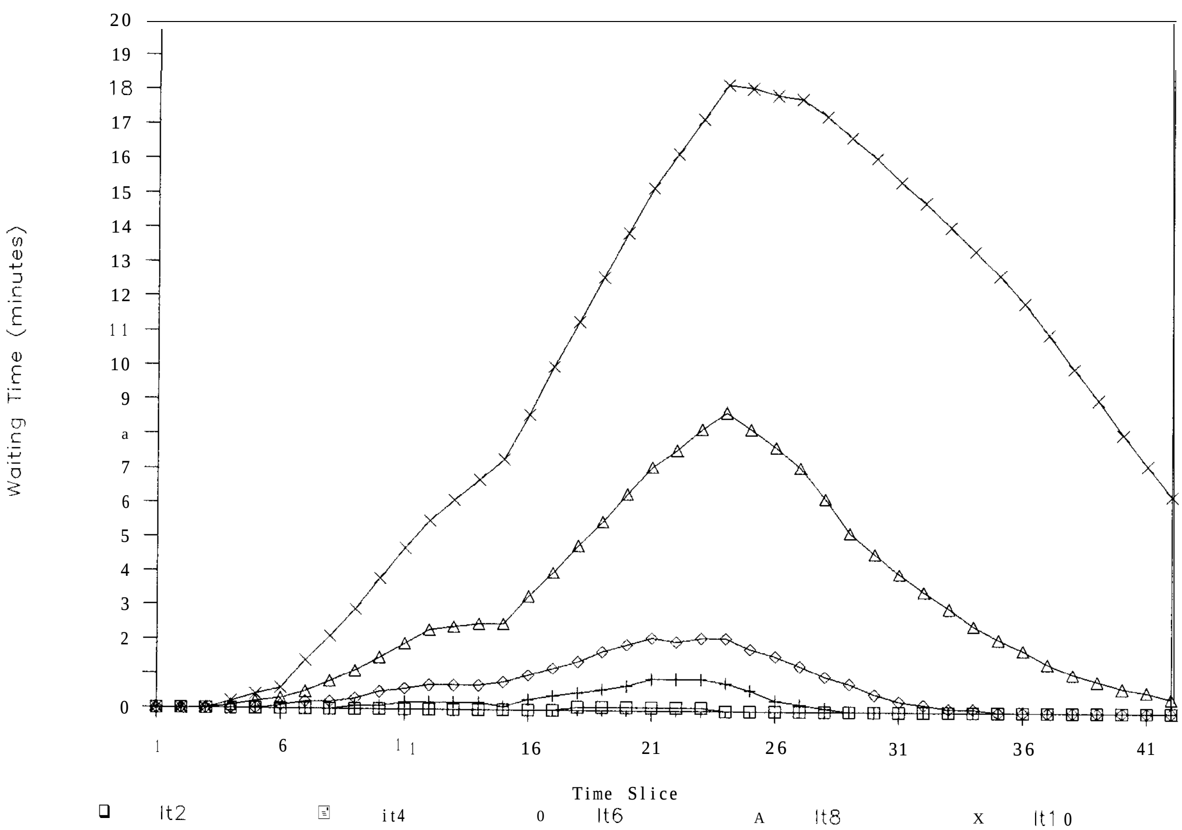

Figure 5. Average time in queue versus arrival time at I-5 with Type 1 travelers, Example 2b (M=.2, b=12, p_1=p_2=.03)

magnitude and the duration of the queue increase. In both cases, the queue duration exceeds two hours in Iteration 10, with a maximum wait exceeding 15 minutes.

A.1 Increased Hazard in Presence of Incidents

Table 3 repeats the set of experiments for I- 5 under the condition $p_2=2p_1$. As should be expected, delay is amplified because the mean length of an incident period has grown. In Example 4b, for instance, delay has increased by 60 to 110%, even though effective capacity has declined by less than 3%. Increasing p_2 greatly increases the 1 ikel ihood of consecut ive incidents, which clearly have an enormous impact on delay. On the other hand, changes in delay tend not to be as great when b=3. If incidents are cleared rapidly, then the roadway is not exposed to the increased hazard of further incidents for very long, so the change in p_2 has only a marginal effect ,

A.2 Delay Threshold

The relationship between delay and capacity can be examined from another perspective : the number of years before delay becomes intolerable. For inst ance, if this threshold is set at a mean delay of five minutes, then the following thresholds result for I-5 ($p_1=p_2$):

Table 3. Time in Queue at I-5 (Minutes)
Type 1 Travelers / p2 = 2.~1

Example 1 : *M = 0.2; p1 = 0.01, p2 = 0.02*

	b = 3		b = 72	
	Avg Q	95%	Avg Q	95%
Iter 2	0	0	0.01	0.11
Iter 4	0.01	0.05	0.1	0.85
Iter 6	0.11	0.65	0.43	2.76
Iter 8	1.44	3.05	2.27	7.45
Iter 10	7.07	9.43	8.73	16.37

Example 2 : *M = 0.2; p1 = 0.03, p2 = 0.06*

	b = 3		b = 12	
	Avg Q	95%	Avg Q	95%
Iter 2	0.01	0.05	0.04	0.19
Iter 4	0.05	0.33	0.38	1.51
Iter 6	0.32	1.46	1.46	5.23
Iter 8	2.1	4.62	4.84	11.87
her 10	8.69	12.26	12.43	21.8

Example 3 : *M = 0.5; p1 = 0.01, p2 = 0.02*

	b = 3		b = 12	
	Avg Q	95%	Avg Q	95%
Iter 2	0.14	1.02	0.96	6.41
Iter 4	0.18	1.15	1.72	11.38
Iter 6	0.59	3.48	2.87	16.74
Iter 8	2.43	7.43	5.63	24.08
Iter 70	8.29	14.61	13.01	34.56

Example 4 : *M = 0.5; p1 = 0.03, p2 = 0.06*

	b = 3		b = 12	
	Avg Q	95%	Avg Q	95%
Iter 2	0.43	2.71	3.83	17.41
Iter 4	0.83	4.37	7.38	25.89
her 6	1.91	7.56	9.83	31.92
Iter 8	4.91	14.76	15.52	40.54
Iter 10	12.86	22.9	24.87	55.48

Example	Iteration	Average Hrly Volume	Effective† Capacity/Hr
1a	10	7726	7970
1b	10	7726	7900
2a	10	7726	7900
2b	9	7358	≈7700
3a	10	7726	7920
3b	9	7358	7740
4a	9	7358	7760
4b	7	6674	≈7250

With the exception of Example 4b, for which the effective capacity is much smaller than other cases, the threshold is always reached in the 9th or 10th Iteration. Overall, we see that small changes in effective capacity produce small improvements in the amount of traffic that can be accommodated. For instance, the 10% capacity increase between Example la and Example 4b facilitates a 15% increase in traffic. This stands in sharp contrast to the earlier prediction that delay is 140% larger for Example 4b. Hence, if continued growth in traffic volume is expected, then small gains in effective capacity may only forestall the date when the threshold is reached by a few years.

B. Deterministic Equivalence

Several examples were re- run as deterministic simulations, using effective capacities as determined from Eqs . 5 and 8. Table 4 applies to the I-5 data for Examples 1, 2 and 4 with $p_2=p_1$ and Examples 1 and 4 with $p_2=2p_1$. In cases where the upper and lower bounds on capacity are appreciably different, results are provided for both cases, with the second-order

† Average of lower bound and upper bound.

Table 4a. Time in Queue at I-5 (Minutes)
Tvpe 1 Travelers / Deterministic Eauivalent / p2 = p1

Example 1 : *M = 0.2, p1 = p2 = 0.01*

	b = 3 Cap = 7968 Avg Q	b = 12 Lb Cap = 7896 Avg Q	 Ub Cap = 7898 Avg Q
Iter 2	0	0	0
Iter 4	0	0	0
Iter 6	0.04	0.09	0.09
Iter 8	1.17	1.52	1.51
Iter 10	6.74	7.6	7.57

Example 2 : *M = 0.2, p l = p2 = 0.03*

	b = 3 Cap = 7904 Avg Q	b = 12 Lb Cap = 7688 Avg Q	 Ub Cap = 7707 Avg Q
Iter 2	0	0	0
Iter 4	0	0	0
Iter 6	0.09	0.28	0.26
Iter 8	1.48	2.69	2.57
Iter 10	7.5	10.18	9.94

Example 4 : *M = 0.5, p l = p2 = 0.03*

	b = 3 Cap = 7760 Avg Q	b = 12 Lb Cap = 7220 Avg Q	 Ub Cap = 7266 Avg Q
Iter 2	0	0	0
Iter 4	0	0.05	0.02
Iter 6	0.21	1.21	0.98
Iter 8	2.25	6.83	6.24
Iter 10	9.27	16.53	15.87

Table 4b. Time in Queue at I-5 (Minutes)
Type 1 Travelers / Deterministic Equivalent / p2 = 2 p1

Example 1 : *M = 0.2, p1 = 0.07, p2 = 0.02*

	b = 3 Cap = 7968 Avg Q	b = 12 Lb Cap = 7889 Avg Q
Iter 2	0	0
Iter 4	0	0
Iter 6	0.04	0.1
Iter 8	1.17	1.55
Iter 10	6.74	7.68

Example 4 : *M = 0.5, pl = 0.03, p2 = 0.06*

	b = 3 Cap =7758 Avg Q	b = 12 Lb Cap =7031 Avg Q	 Ub Cap = 7218 Avg Q
Iter 2	0	0	0
Iter 4	0	0.22	0.05
Iter 6	0.23	2.3	1.22
Iter 8	2.34	9.37	6.86
Iter 10	9.46	19.33	16.56

approximation used for the upper bound. In most cases, the bounds are very close, and either result would suffice.

There was reasonably close agreement between the results, especially in the later iterations when queues are large. The largest difference between the deterministic and stochastic runs occurs toward the end of the queue in early iterations, as is revealed by comparing Figure 6 (deterministic, Example lb, I- 5, $p_1=p_2$) to Figure 4 (stochastic). The waiting time curve drops much more gradually in the stochastic case than in the deterministic case. This can be attributed to the occasional incident that creates a residual queue lasting well beyond the time the queue ordinarily vanishes. The stochastic simulations also exhibit much less predictable waits. This is revealed in Tables l-3 by comparing the 95th percentiles of the waiting time distribution. It is also revealed in Figure 7, which plots the 95th percentile for Example lb of Table 1. In many cases, the 95th percentile is several times larger than the mean, especially in earlier iterations when queues are ordinarily small.

Overall, it seems that a deterministic simulation, with the same effective capacity, can provide useful results. Moreover, it seems that when there is little surplus capacity, the expected effect of incidents is similar to that of reducing a deterministic capacity by a few percent (however, deterministic simulations tend to underestimate delay when there is surplus capacity). This holds true even when the majority of delay is caused by incidents because the marginal effect of a small capacity change can be quite large.

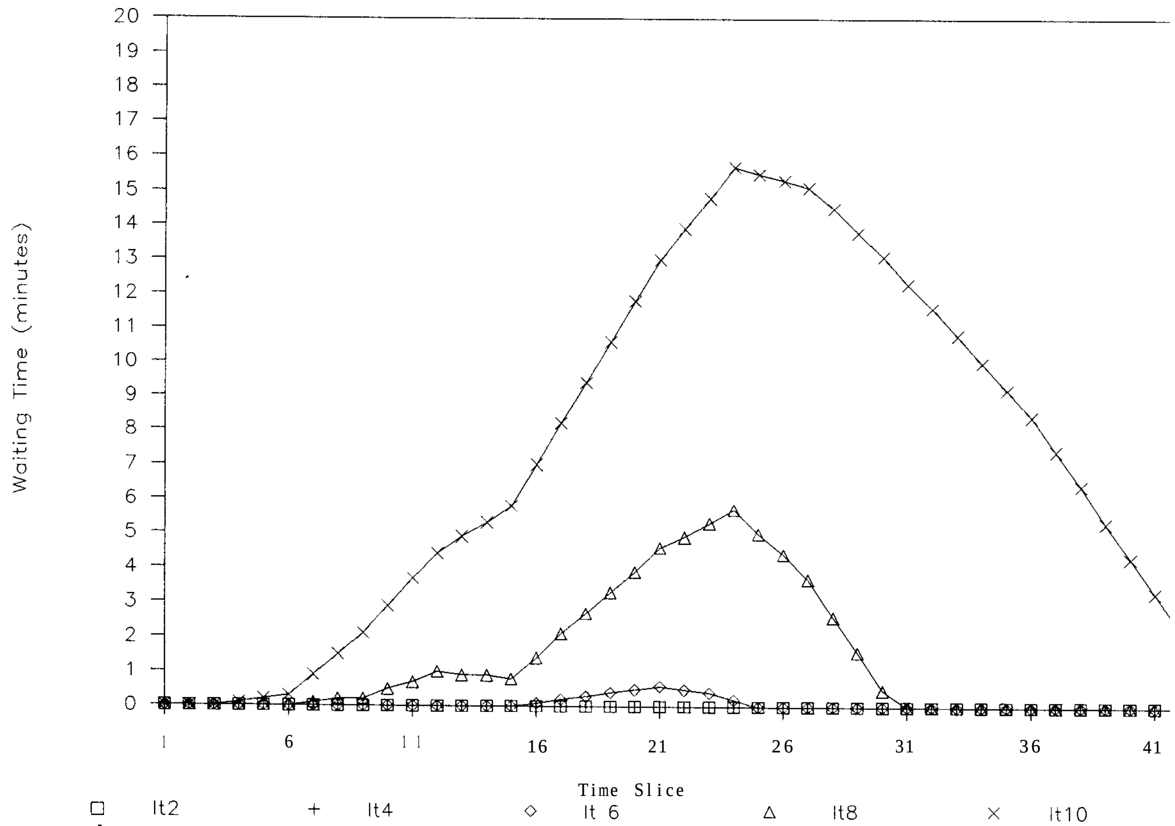

Figure 6. Average time in queue versus arrival time at I-5 with Type 1 travelers. Deterministic simulation of Example 1b.

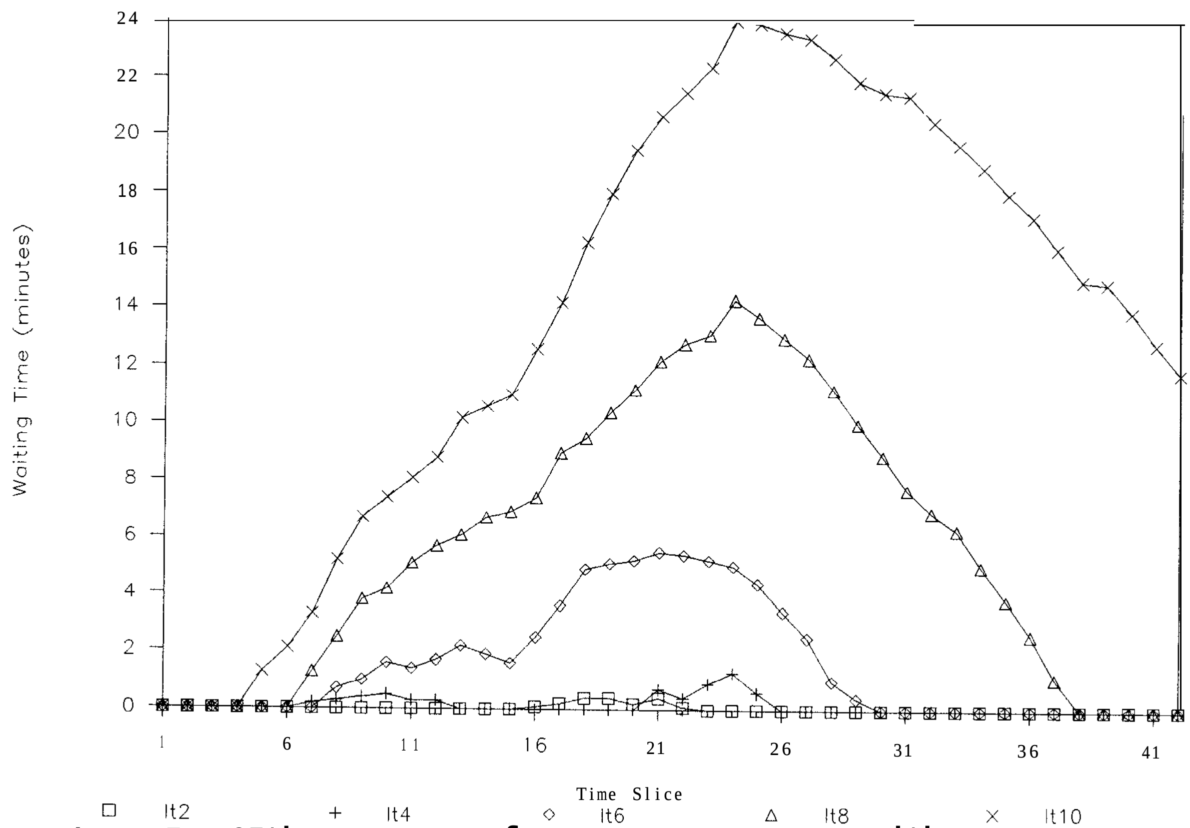

Figure 7. 95th percentile of time in queue at I-5 with Type 1 travelers Example lb (M=.2, b=12, $p_1 = p_2 = .01$)

C. Changes in Arrival Pattern

A third set of experiments was designed to measure the effects of changes in vehicle arrival times. In these experiments, 100% of the travelers fell in the Type 3 category, and all scenarios were drawn from I-5, Example lb ($M=.2$, $p_1=p_2=.01$, $b=12$). Experiments were run for different travel costs:

I.	Lateness = \$48/hour	Travel = \$12/hour	Earliness = \$3/hour
II.	Lateness = \$18/hour	Travel = \$12/hour	Earliness = \$8/hour
III.	Lateness = \$12/hour	Travel = \$12/hour	Earliness = \$12/hour

The first case would apply if travelers must arrive at their destinations by precise deadlines, such as the start of work. Consequently, they may leave home early to insure that they are not late. In the last case, travelers have some flexibility in selecting their arrival times. They may shift their arrival times either before or after their preferred times if cost is reduced.

Table 5 provides results for all three cases. Compared to Table la, Type 3 travelers experience much longer delays, especially in Case I. The cause is revealed in Figure 8a (Case I) . Because Type 3 travelers want to arrive by set times, a form of dysfunctional competition ensues. Travelers shift to earlier arrival times to ensure that they reach their destination on time. This creates a sharp peak in the arrival pattern that imposes added delay. In the other extreme, Case III (Figure 8b) , travelers still shift to earlier arrival times, though the shift is less pronounced. In both cases, shifts accentuate delays. Nevertheless, with suitable earliness and lateness costs, it is possible for the arrival pattern to become more spread out in later iterations, in which case behavioral changes will cause delays to decline.

Table 5. Time in Queue at I-5 (Minutes)
Type 3 Travelers / No Reneaing

(p1 = p2 = 0.01, M = 0.2, b = 72)

Type 3 Cost 48 /12/ 3

	Total Arrivals	Avg C?	95%
Iter 2	17939	0.01	0.05
Iter 4	19778	0.04	0.13
Iter 6	21805	0.68	2.39
Iter 8	24040	9.61	14.22
Iter 10	26504	40.42	47

Type 3 Cost 18 /12/ 8

	Total Arrivals	Avg Q	95%
Iter 2	17939	0.01	0.05
Iter 4	19778	0.04	0.13
Iter 6	21805	0.26	1.95
Iter 8	24040	3.04	7.53
Iter 10	26504	16.44	23.11

Type 3 Cost 12 /12/ 12

	Total Arrivals	Avg Q	95%
Iter 2	17939	0.01	0.05
Iter 4	19778	0.04	0.13
Iter 6	21805	0.26	1.95
Iter 8	24040	2.45	6.95
Iter 10	26504	14.46	21.18

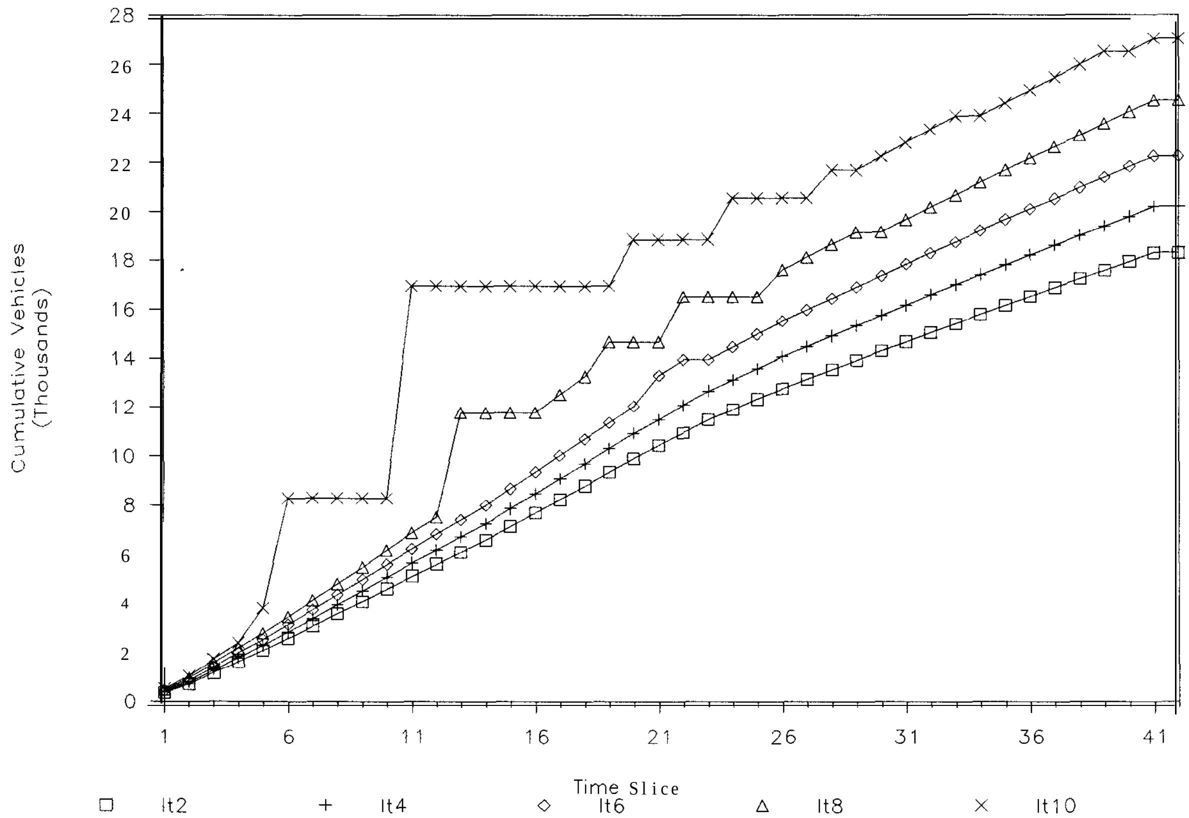

Figure 8a. Cumulative arrivals for Type 3 travelers without reneging (Example lb, traveler costs of 48/12/3).

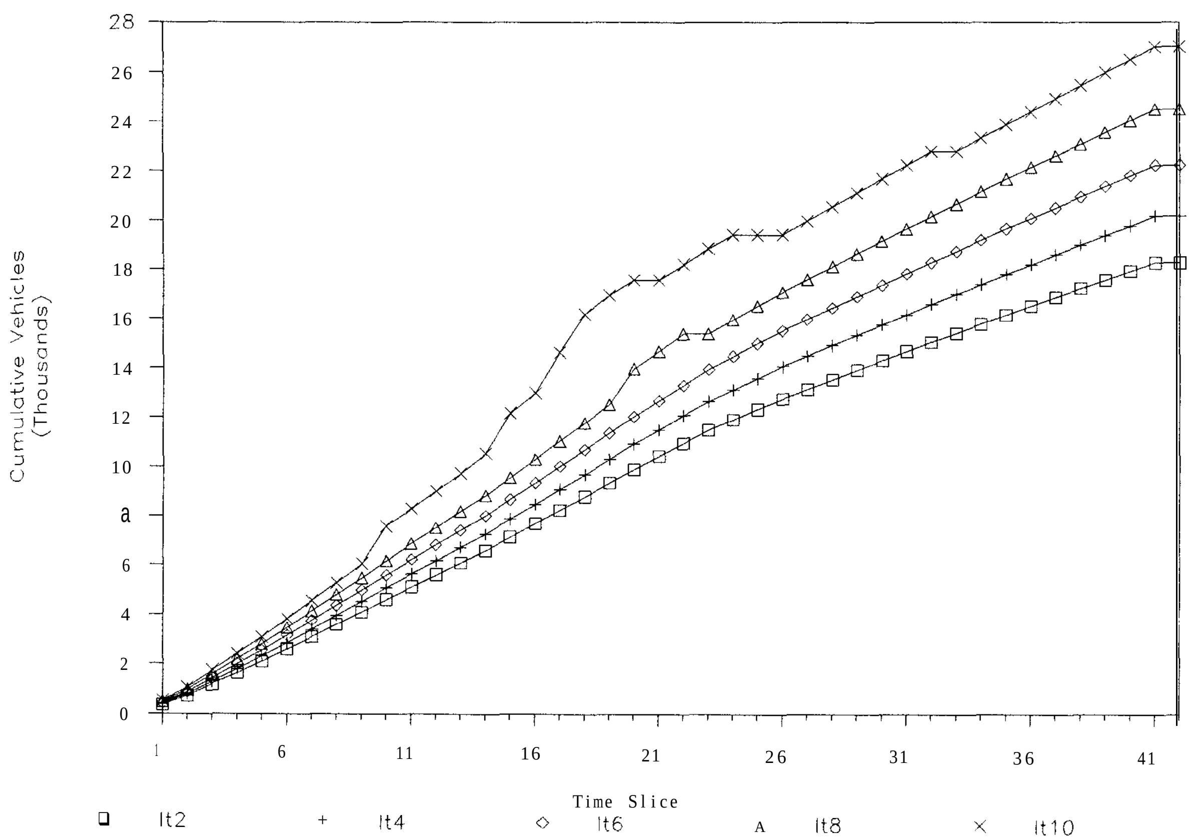

Figure 8b. Cumulative arrivals for Type 3 travelers without reneging (Example lb, traveler costs of 12/12/12).

D. Reneging

In the final set of experiments, all travelers again were Type 3, but they were now allowed to renege. A base reneging cost function was specified, with the following form:

$$R = \left[\frac{\min(C,\beta)}{\beta}\right]^{\alpha}, \quad C \leq \beta , \tag{13}$$

where R is the proportion of travelers who would renege when the cost of traveling equals C. Figure 9 illustrates the structure of the function. In experiments , Eq. 13 was discretized into four equal-sized regions to match data requirements of BTS.

Simulations were run for the cost functions of Cases I and III of Section C, again under Example lb for I- 5. The following parameters were used in the reneging function:

Type la: $\beta = 6, \quad \alpha = 1$	Type lb: $\beta = 6, \quad \alpha = 2$	Type lc: $\beta = 6, \quad \alpha = 4$
Type 2a: $\beta = 12, \alpha = 1$	Type 2b: $\beta = 12, \alpha = 2$	Type 2c: $\beta = 12, \alpha = 4$

Type la creates the most reneges, especially for small travel costs, while Type 2c creates the fewest reneges. Type 1 results are found in Table 6 and Type 2 results are found in Table 7.

Comparing results to Table 5, reneging clearly has a moderating effect on waiting. Delays do not increase nearly as rapidly, especially when a = 1 and β = 6. In these cases, traffic growth slows considerably after the sixth iteration, when measurable queues first materialize. Nevertheless, reneging

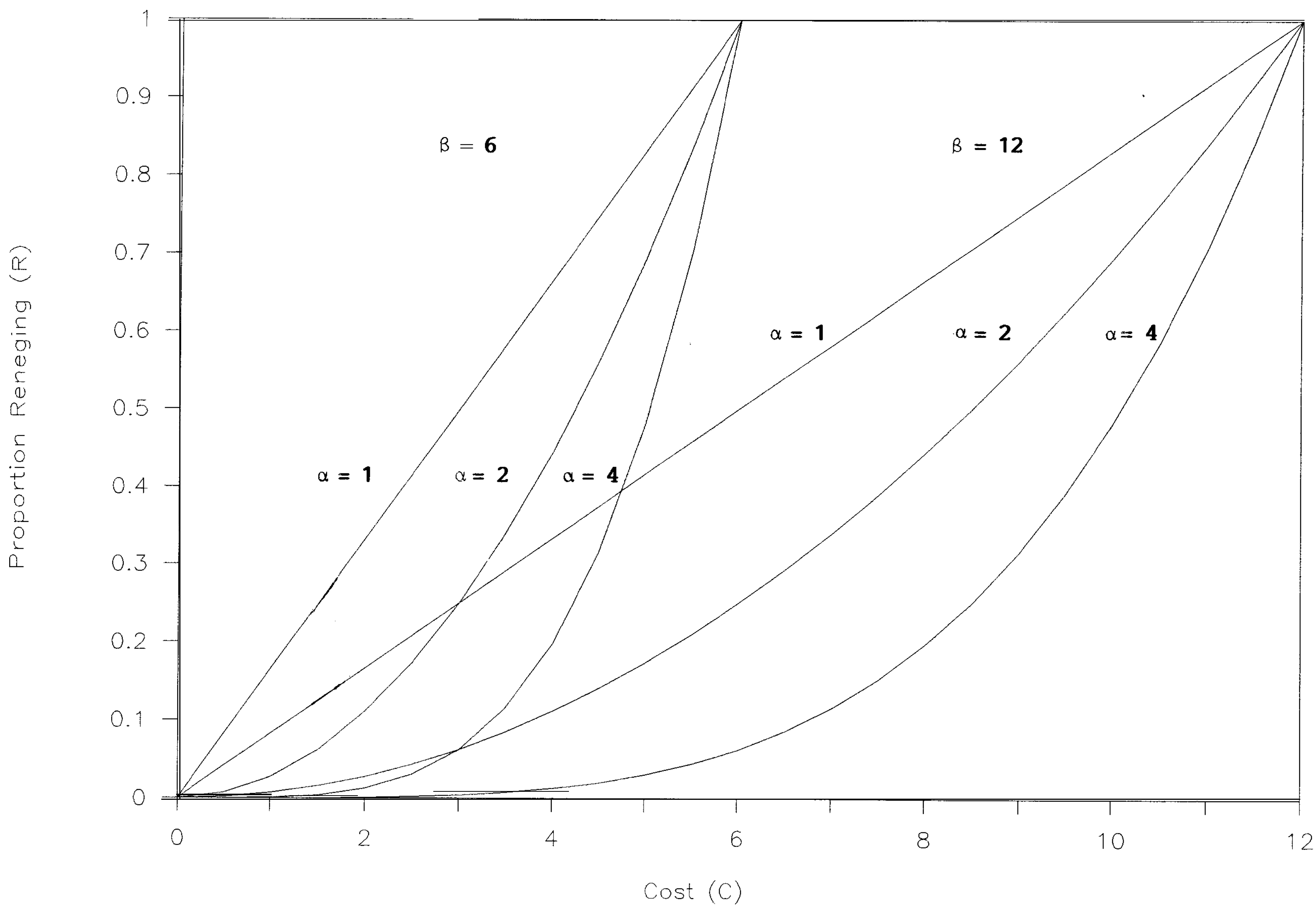

Figure 9. Reneging function used in simulations.

Table 6. Time in Queue at I-5 (Minutes)
Type 3 Travelers / with Reneging $\mathcal{B} = 6$

(pl = p2 = 0.01, b= 12, M = 0.2)

Alpha.= 1		48/1 2/3			12/12/12	
	Total Arri	Avg Q	95%	Total Arri	Avg Q	95%
Iter 2	18301	0.01	0.05	18301	0.01	0.05
Iter 4	20177	0.04	0.13	20177	0.04	0.13
Iter 6	22246	0.67	2.34	22246	0.25	1.91
Iter 8	22633	4.89	7.44	23113	0.75	3.23
Iter 10	23410	8.83	11.33	24159	1.83	4.55

Alpha =2		48/12/3			12/12/12	
	Total Arri	Avg Q	95%	Total Arri	Avg Q	95%
Iter 2	18301	0.01	0.05	18301	0.01	0.05
Iter 4	20177	0.04	0.13	20177	0.04	0.13
Iter 6	22246	0.67	2.43	22246	0.25	1.91
Iter 8	24029	8	11.97	24187	1.78	5.73
Iter 10	24070	19.93	23.68	25451	6.74	11.53

Alpha = 4		48/12/3			12/12/12	
	Total Arri	Avg Q	95%	Total Arri	Avg Q	95%
Iter 2	18301	0.01	0.05	18301	0.01	0.05
Iter 4	20177	0.04	0.13	20177	0.04	0.13
Iter 6	22246	0.67	2.43	22246	0.25	1.91
Iter 8	24526	9.42	13.94	24526	2.41	6.82
Iter 10	25291	31.16	37.15	26002	10.77	16.33

Table 7. Time in Queue at I-5 (Minutes)
Type 3 Travelers /with Reneaina $\mathcal{B}$ = 12

(pl = p2 = 0.01, b= 12, M = 0.2)

Alpha = 1		48/1 2/3			12/12/12	
	Total Arri	Avg Q	95%	Total Arri	Avg Q	95%
Iter 2	18301	0.01	0.05	18301	0.01	0.05
Iter 4	20177	0.04	0.13	20177	0.04	0.13
Iter 6	22246	0.67	2.34	22246	0.25	1.91
Iter 8	24526	9.42	13.94	24526	2.41	6.82
Iter 10	24916	23.33	28.32	25228	8.1	12.59

Alpha = 2		48/12/3			12/12/12	
	Total Arri	Avg Q	95%	Total Arri	Avg Q	95%
Iter 2	18301	0.01	0.05	18301	0.01	0.05
Iter 4	20177	0.04	0.13	20177	0.04	0.13
Iter 6	22246	0.67	2.43	22246	0.25	1.91
Iter 8	24526	9.42	13.94	24526	2.41	6.82
Iter 10	26390	36	42.63	26571	12.57	18.58

Alpha = 4		48/1 2/3			12/12/12	
	Total Arri	Avg Q	95%	Total Arri	Avg Q	95%
Iter 2	18301	0.01	0.05	18301	0.01	0.05
Iter 4	20177	0.04	0.13	20177	0.04	0.13
Iter 6	22246	0.67	2.43	22246	0.25	1.91
Iter 8	24526	9.42	13.94	24526	2.41	6.82
Iter 10	27040	39.82	46.46	27040	14.46	20.63

does not eliminate peaks in the arrival pattern. In Figures 10a and 10b (Case I costs of 48/12/3), growth in delay between iterations is more due to arrival time shifts than to traffic growth. This is less true for Case III (Figures 11a and 11b). Nevertheless, peaking in the arrival pattern still plays a significant part in queue growth.

In some simulation runs (not shown) we even found that total traffic volume could decline in later iterations. This may seem paradoxical, but there is a simple explanation. As time progresses, the peak in the arrival pattern becomes more and more pronounced. Because peaked arrivals lead to more delay, travel costs are driven up and more people choose not to travel through the corridor. If the peak becomes even more accentuated, delay may continue to rise even after total traffic has declined. On the other hand, if the arrival pattern becomes spread out, then more vehicles could travel without incurring large delays. This is one of the motivations for congestion pricing. By reducing traffic levels during the peak, residual delays are reduced and more vehicles will find it cost-effective to travel after the peak in arrivals.

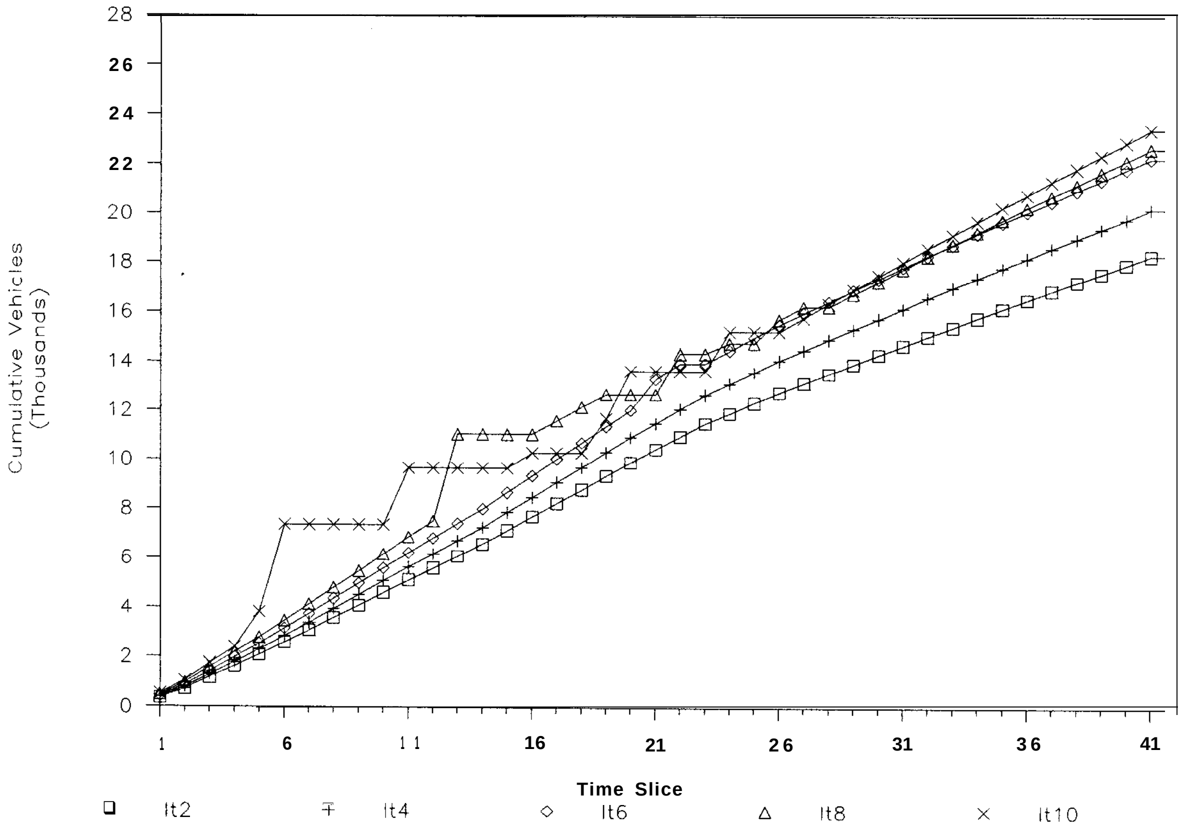

Figure 10a. Cumulative arrivals for Type 3 travelers with reneging, alpha=1, beta=6, Example lb, traveler costs of 48/12/3

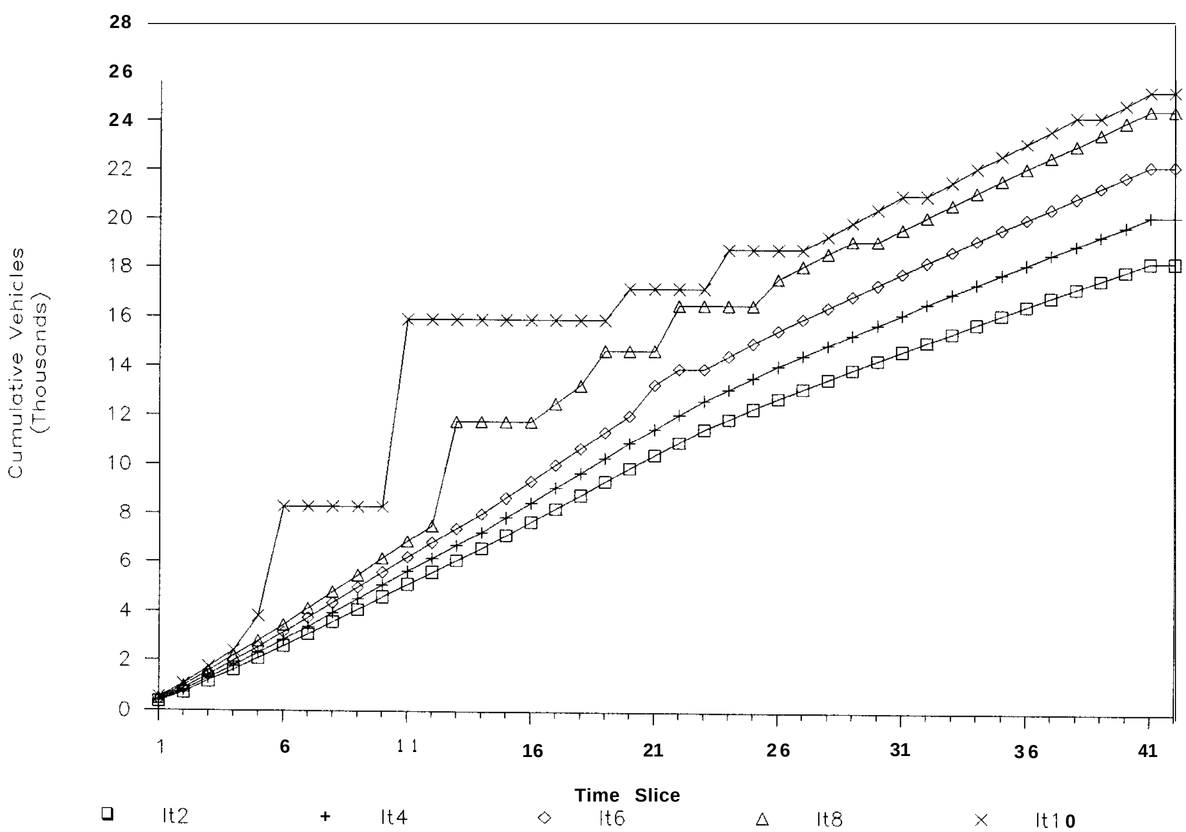

Figure 10b. Cumulative arrivals for Type 3 travelers with reneging, alpha=4, beta=6, Example lb, traveler costs of 48/12/3

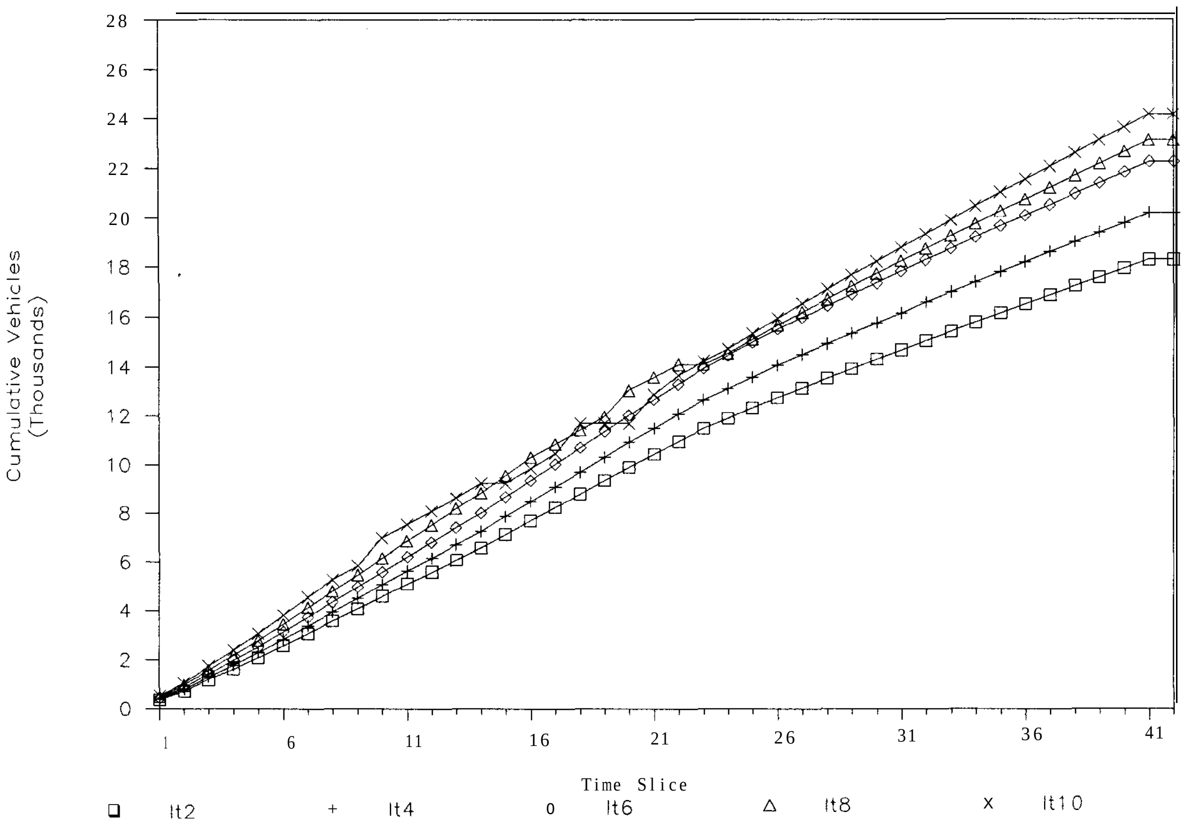

Figure 11a. **Cumulative arrivals for Type 3 travelers with reneging alpha=1, beta=6, Example 1b, traveler costs of 12/12/12.**

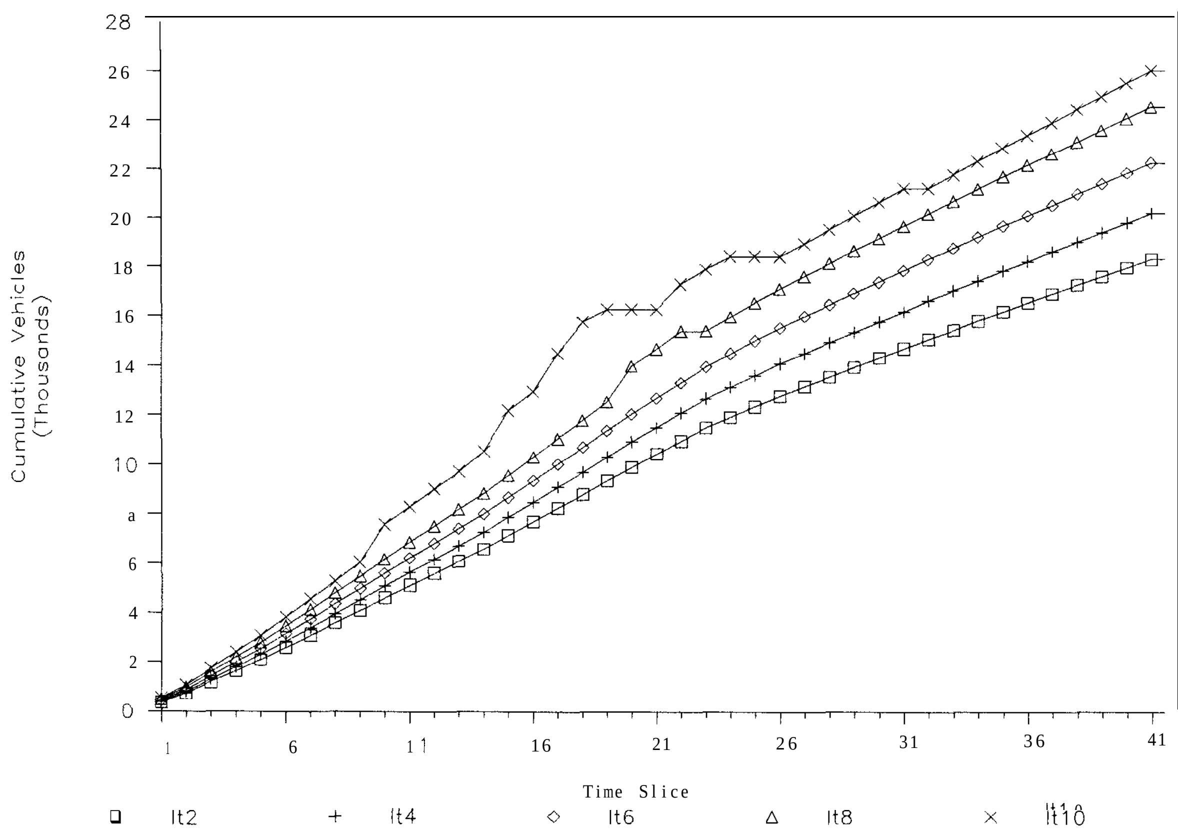

Figure 11b. **Cumulative arrivals for Type 3 travelers with reneging alpha=4, beta=6, Example 1b, traveler costs of 12/12/12.**

CHAPTER 5. CONCLUSIONS

BTS provides a framework for analyzing highway performance in the presence of incidents and changes in traveler behavior. As demonstrated in this report, both factors have substantial effects on highway performance. Our findings include:

- The presence of nfrequent incidents can, under some scenarios, cause large percentage increases in vehicle delay.
- Delay is highly sensitive to arrival time behavior. If travelers have rigid deadlines, peaks in traffic volume are accentuated and delays can become very large. If travelers have flexible deadlines then peaking is much less pronounced.
- Reneging has a moderating effect on delay, especially when base traffic has become large. Failure to account for reneging will cause delay to be overestimated.
- In some instances, deterministic simulations, with comparable effective capacity, can substitute for the more time consuming stochastic simulations.

In the absence of behavioral changes, even small improvements in highway performance translate into significant changes in delay. In reality, any improvement in performance will likely lead to increased traffic volume, as fewer travelers renege. Consequently, reductions in delay may be less than predicted under a static model. For this reason, extra caution must be exercised with respect to estimating the benefits of incremental improvements in highway performance, such as incident management or traveler information strategies. It may be tempting to predict that increases in effective capacity on the order of 2-5% will lead to reductions in travel time of 50% or more. In reality, the consequence of a 2- 5% capacity gain may be that 2-5%

more travelers are able to travel during the peak, with only marginal travel time reductions.

Overall, this study demonstrates the inherent difficulty in predicting the benefits of new highway technologies. Most importantly, it should be a routine practice to measure highway performance (in the form of traffic flows and travel times) both before and after every significant highway project. In this way, assessments of future highway investments can be placed on an objective footing.

There is also a need for empirical research on highway incidents. Limited data is available on their effects, frequency and magnitude. Without this data, it is difficult to objectively formulate strategies for their elimination.

Secondarily, future research could be invested in extending the capabilities of BTS. For instance, BTS could be revised to allow for simulation of real-time traveler response to incidents. BTS also could be expanded for simulation of network wide impacts of incidents.

REFERENCES

Hall, R.W. (1991) "Travel Time Benefits of New Transportation Technologies: The Case of Highway Automation," PATH Working Paper 91-4.

Lin, W. and Hall, R.W. (1991) "Bottleneck Traffic Simulator (BTS) Version 1.1)" PATH Working Paper 91-7.

ACKNOWLEDGEMENTS

Our appreciation goes to Wei Lin for developing the program BTS and providing his careful comments on this paper.