SOME IDEAS FOR FREEWAY CONGESTION MITIGATION

WITH ADVANCED TECHNOLOGIES

Carlos F. Daganzo, Jorge Laval and Juan Carlos Muñoz
Institute of Transportation Studies, University of California, Berkeley
(June, 2001)
(Revised, December 2001)

ABSTRACT

This report presents twelve ideas for improving freeway performance that have become feasible with the advent of advanced technologies. The strategies have in common that they can be rigorously tested. Their measures of performance can be reliably obtained and do not depend on the accuracy of data-hungry, large-scale models.

1   INTRODUCTION

Recent experimental work (Muñoz and Daganzo, 2000) has established the serious deleterious effects of first-in-first-out (FIFO) queues on freeways. FIFO queues are very common, and very damaging. They correspond to what some researchers call "synchronized flow" or "1-pipe flow"-- a state in which all lanes flow at the same speed, with few lane-changes; see Daganzo, (1999) and Zhang (2001). Although FIFO queues can be eliminated by severe ramp metering (e.g. as in Minneapolis, Minnesota), indiscriminate actions of this type can reduce freeway flow through key bottlenecks, transferring the congestion problem to the on-ramps and local street network. Society can end up being much worse off with these kinds of actions, despite an improvement in freeway speeds. This is not widely recognized in some circles, where metering is viewed as a panacea, but it should be obvious; after all, in addition to high speeds, freeways have the ability to store large amounts of vehicles, and this resource should not be wasted cavalierly. It can actually be very beneficial. This report is a step in this direction.

It is well known that non-FIFO queues (with different speeds on different lanes) have a much less severe effect on the traffic stream than FIFO queues. Therefore, freeway congestion could be reduced without transferring all the queues to the surface streets, if freeway drivers could be induced to accept non-FIFO queues. The empirical observations in Muñoz and Daganzo (2000) establish that this is indeed possible. This reference identified a stable non-FIFO queue, in which only some of the freeway lanes were congested, and showed that drivers accepted this situation for a long time. Related experimental work, also described in that reference, demonstrates that changing the character of a queue from FIFO to non-FIFO could improve by 50% or more the capacity of some bottlenecks. The character of a freeway queue, however, can only be changed through careful application of innovative control methods that require "advanced" technologies.

This report presents a list of twelve capacity-expanding control strategies that, like the above, can only succeed with some form of advanced technology. Some of the proposed methods would change the character of queues as suggested above, and others would not. However, all the strategies are more refined (and potentially more effective) than the simple form of ramp metering mentioned at the outset. Although some of the methods may seem rather unconventional, they are easy to test.

In all cases, benefits can be quantified accurately from field measurements without modeling. A common goal is to reduce total delay, summed over all users. The strategies are grouped into the following four sets for ease of explanation:

(a) Lane assignment actions,

(b) HOV actions,

(c) Ramp metering actions, and

(d) Miscellaneous control actions.

Sections 2 through 5 explain each of the above sets, and section 6 presents some conclusions.

2 LANE ASSIGNMENT ACTIONS

The goal of these strategies is reducing the "friction" of lane-changing maneuvers by clearly defining weaving sections for problematic origin-destination (O-D) pairs. Currently, weaving sections are spontaneously defined by drivers' lane-changing decisions, and this often causes undesirable jams. Variable message signs (VMS) and associated technologies for monitoring driver behavior can be used to segregate drivers by destination and force them to change lanes only where allowed.

Frequently, freeway bottlenecks create queues that entrap vehicles not headed for the bottleneck, as shown in Fig. 1. For example, Muñoz and Daganzo (2000) report that a queue that spilled back from a congested off-ramp induced FIFO (or 1-pipe) behavior in all 5 lanes of its freeway for over an hour. It was observed that the capacity for through vehicles (those not causing the problem) was reduced by the FIFO effect to 1,500 vphpl; a 25% reduction. This could be avoided by breaking the FIFO behavior.

FIFO can be broken by VMS strategies that would tell drivers in what lane to be according to their destination. Messages would be similar in spirit to pavement markings but programmed for fixed time periods or, even better, traffic actuated. These lane-assignment strategies should be strengthened by signs and regulations banning last minute lane changes to "cut in" a queue. When a vehicle "cuts in" a queue it has the following effects on flow: First, its deceleration maneuver induces a temporary queue in the vehicle's original lane if the flow is high. (With enough repetitions,

this effect will create a permanent FIFO queue across all lanes.) Second, if the lane-changing vehicle forces its way into its destination lane, the vehicles in this lane may have to decelerate suddenly. As a result, the discharge flow across both lanes, downstream of the lane-change, is reduced. (This is the observable result we called "friction".) Note too, that sudden lane-changing maneuvers create a safety hazard.

In cases where a queue spills over from an off ramp, as in Fig. 1, forcing exiting traffic to join the queue on the shoulder lanes from the back would enhance freeway capacity. This can be accomplished by devoting some shoulder lanes to the queue, as shown in the figure. One must decide (with limited information) both, the length and the number of lanes that should be assigned to the congested exit movement. Ideally, these decisions should be updated continuously over time.

The decisions involve a delicate balance. On the one hand, there is a benefit to assigning many lanes to the exit queue because this shortens it, and prevents it from affecting upstream ramps. On the other hand, if too many lanes are allocated to the congested movement, the exit queue will leave less room for the remaining (through) vehicles, and this could activate a new FIFO bottleneck. It would completely negate the benefits of the control measure.[1] Note as well that the number of lanes can be increased but cannot be decreased easily once the queue has formed.

For smooth flow, weaving maneuvers should take place upstream of the queue, and not alongside it, as shown in the figure. Therefore, the VMS's should span the length of the queue, and probably extend at least ¼ mile upstream the back of the queue. This requires sensing the back of the queue and actuating the VMS's accordingly. The VMS's upstream of the queue would encourage lane-changes. The VMS's alongside the queue would ban last minute lane changes to "cut in" the queue.

In simple cases, such as the one in Figure 1, common sense can lead us to an optimum design, i.e. an optimum location of weaving segments, sensors, etc. In a more complicated situation with several origin-destination (O-D) pairs, a simple simulation model could be useful.

Site-specific complications can arise. If the nearest upstream ramp is close, the exit queue may eventually reach it. If it is an on-ramp, traffic from the ramp should be allowed to join the

---

[1] The theory for these types of bottlenecks has been examined in Daganzo (1997).

queue or pass through it. If it is an off-ramp, traffic for this destination should be encouraged to join the back of the queue. Clearly, the location of the back of the queue dictates the messages that should be posted. Sensing devices and VMS's should be placed and programmed accordingly.

If the through flow exceeds the capacity of the through lanes (even when only one lane is allocated to the exit queue) then the proposed strategy will not work. In this case, one should strive for increasing the capacity of the off-ramp. If this capacity is restricted by a traffic signal, or the conditions on the local streets, then the traffic signal and/or the city streets should be operated to help the freeway, assuming this is fair and politically feasible.

Perhaps, lane assignment strategies can also be used to confine a freeway queue to the left lanes. This is more questionable, but it would allow short trips that take place entirely upstream of a bottleneck to use the freeway unhindered. Vehicles would not be allowed to join the queue after traveling on the freeway for more than (say) ½ mile. This would require signs, and a way of enforcing the trip length restriction; e.g., by way of fines. The main technological problem is keeping track of cheating vehicles.

Lane-changing restrictions can also be useful in other instances. For example, when "last-minute" lane-changes just upstream of a bottleneck reduce its capacity (as may happen at uphill lane-drops and sag bottlenecks) lane-changes could be banned near the bottleneck and encouraged further upstream where their effect would be less pernicious.

3 HOV ACTIONS

An important way to increase capacity involves HOV lanes. If they cut through a bottleneck carrying less than capacity flow when the other lanes are queued, then capacity is lost. The effects upstream of the bottleneck can be disastrous. Fortunately, there are at least two technology-based ways in which the capacity of an HOV lane can be recovered without hindering HOV's significantly. One of these approaches, "coercive turn-taking", was described in Daganzo (2000) and will be outlined in section 5.2. The other approach is described here.

Suppose that an HOV lane passing through a bottleneck is under-utilized. Can we recover the lost capacity without penalizing HOV's? It is proposed to activate and deactivate dynamically the HOV designation in a longitudinal section of the freeway that includes the bottleneck in its entirety.

During off-periods all vehicles could use the section; during on-periods only HOV's. Variable message signs would convey the message. The HOV lane would always be active upstream of the dynamic section. Speed sensors would detect the presence of a queue in the HOV lane upstream of the dynamic section. This information would guide the control actions to prevent this queue from growing too long. Figure 2 shows the structure of such a system and how it should be operated.

The idea is to ensure that HOV's receive priority in passing through the bottleneck while ensuring that the HOV lane is fully utilized (discharging at capacity). Figure 3 shows that this can be accomplished in two phases, mimicking an actuated traffic signal, as follows:

1. The HOV designation of the dynamic section is turned off to increase the bottleneck flow, until a queue of HOV's is detected upstream. This is shown in parts (a) and (b) of the figure. During this "green" phase the HOV lane would carry capacity flow because non-HOV's can merge into it, as indicated by the black arrows of the figure. Because of this interference, we assume that there is a maximum flow of HOV vehicles, q_c, that can enter the dynamic HOV section during the green periods.

2. Then, the VMS restriction would be turned back on, as shown in part (c) of the figure. During this "red" phase the HOV lane would carry capacity flow because the HOV queue is discharging as it dissipates, but the HOV lane flow would eventually decline, after dissipation.

3. The VMS should then be turned off again, and the (a), (b), (c) cycle should be repeated.

The method may or may not work, depending on compliance, enforcement and human-factors issues, but it would be relatively easy to test.

Note that if the flow of HOV's is smaller than q_c then a queue of HOV's will never grow in Step 1. In this case, the dynamic HOV section would be off all the time (open to everybody) but HOV's would not be hindered significantly.

Figure 3 shows a time-space diagram for the trajectories of HOV vehicles, upstream of the dynamic section. This section is of length g and the flow of HOV's is f. As soon as the queue reaches a critical length (d on the figure), the dynamic HOV designation is turned on, allowing the front of the queue to move upstream until reaching the back of the queue, as shown in the figure. We can see from the geometry of the figure that the maximum queuing delay to HOV vehicles is linearly

related to this critical length. The trajectories on the upper middle part of the figure are not depicted because there is no room here to discuss the behavior of traffic in the dynamic section itself, although models can be easily developed. These models can be used to determine the length of the dynamic HOV section. The analogy with actuated traffic signals provides valuable insight about the operational parameters that may lead to an efficient performance. We think this is a topic that deserves further investigation.

Technology: To implement this strategy, one needs detectors to sense when the back of the queue has reached the location that triggers the on-phase, VMS's to display the HOV status, and (in the future) variable pavement markings to reinforce the message. Upstream loop-detectors could help to anticipate future flows, and improve the control algorithm.

4 RAMP-METERING ACTIONS

4.1 Destination-Specific Metering at On-Ramps

When the destination pattern at an on-ramp is such that only a few vehicles pass through a downstream bottleneck, but most do not, it may be advantageous to meter both traffic types differently. Obviously, this strategy cannot be applied to single-lane ramps. However, at many one-lane sites it is possible to "squeeze" another lane in. Lower speeds should be enforced at such sites to prevent safety hazards.

Traditional ramp-metering can be viewed as a queuing system with one server (working at the rate of the meter), two types of customers (drivers who will and drivers who will not pass through the bottleneck) and a common FIFO queue. Some flexibility can be gained by splitting this queue in two, by customer type, and metering the two queues separately. As shown in Figure 4, each queue could be assigned to different on-ramp lanes. These lanes could be metered dynamically. In the particular case of the figure, where the bottleneck is on the "main-line", one could allow exiting traffic to proceed unhindered, while metering the through flow below the capacity of the bottleneck. If there is only one bottleneck, as in our example, then determining the optimal metering rates at a system of on-ramps is a simple optimization problem; see Lovell and Daganzo (2000). Otherwise, the optimization problem is more complex; Erera et al (2001).

Another possible application of non-FIFO metering is when the bottleneck is at an exit ramp. In that case one would only meter vehicles destined for the congested ramp. The benefits of this strategy may not be fully realized if only some of the ramps can be metered by destination because severely affected drivers at these ramps could then switch to the other (FIFO) ramps and congest the off-ramp. Furthermore, vehicles not going through the bottleneck might prefer the non-FIFO ramps, where they are sure to find no queues. This jockeying will change the O-D pattern and may reduce the benefits.

Technology: In all cases, technology is needed to track and penalize uncooperative drivers who take the fast lane but pass through the bottleneck. This can be done in a variety of ways, some less intrusive than others, and should be investigated in future work.

4.2 Dynamic Off-Ramp Management

This strategy is suitable for freeways that carry an unusual high demand for some off-ramps over a relatively short period of time; e.g., a special event such as a football game in a nearby stadium (Figure 5). In these cases, the exit queues may back up, form a FIFO queue that entraps non-special event traffic, and cause great system-wide congestion.

If some neighboring off-ramps are not sufficiently used, congestion can be mitigated by selectively closing the congested ones and diverting traffic to those less used. With the right geometry, closing congested off-ramps for short periods of time (e.g., a few minutes at a time) could be very advantageous for the system as a whole, even though it may delay some fans a little more. Strategies of this type would serve the exiting traffic rapidly, by as many off-ramps as possible, taking advantage of the freeway and the off-ramps ability to store vehicles. This can be done in two ways: giving and not giving drivers optional advice.

Method 1: Direct control without advice to improve usage of downstream off-ramps. This can be simply achieved by denying access to the preferred off-ramp once the off-ramp queue has emerged onto the freeway and the precursor signs of FIFO are detected. With access denied, the queued vehicles on the freeway will be forced to drive further downstream and use, perhaps even queue at, another off-ramp. The freeway bottleneck may discharge at maximum rate during this period. When the preferred off-ramp is nearly empty, it can be reopened. A queue may then again grow somewhere within the off-ramp and later emerge once again on the freeway. Because the

bottleneck discharge rate will remain high until FIFO conditions develop, the off-ramp does not need to be closed until they are imminent. Experience shows that it takes many minutes between the emergence of the off-ramp queue onto the freeway and the onset of FIFO. Of course, the off-ramp closure should be coordinated with the status of the downstream off-ramp(s).

This simple mechanism ensures that through traffic always flows freely because a FIFO queue is never allowed to form on the freeway. It also benefits exiting traffic because it is served at a combined higher rate. Note that ramps can be closed physically (e.g., with police cars) without VMS's. Thus, a demonstration project should be easy to carry out.

Method 2: Use of VMS's to improve utilization of upstream off-ramps. Method 1 can be enhanced by also redirecting traffic to upstream ramps. Variable message signs can be used to do this, but forceful messages may be needed because an important proportion of users may not heed simple advice. To improve the reliability of the control, one could try to broadcast expected delays on various ramps, or even advertise that certain ramps are temporarily closed.

If a precise method for redirecting traffic can be devised, then the system can be optimized. Laval and Munoz (2001) develops optimality conditions for the version of this problem in which the off-ramp bottlenecks are at the city streets, the off-ramps can store a significant number of vehicles and queues are not allowed on the freeway. Let T_i be the extra circuity time associated with ramp i ($i= 1, 2…$) and rank all the ramps in order of increasing T_i (with $T_1 = 0$). It is shown in Laval and Munoz (2001) that a single surge in exits is optimally managed if:

1. As the exit flow increases, off-ramps enter into service in order of increasing i. On-ramp i does so when the total exit flow first equals the combined capacity of ramps 1 to i. When an off ramp enters service it receives a flow greater or equal to its capacity and is immediately saturated. It then remains at capacity until the time identified under condition 2.

2. As the combined exit flow decreases, it should be allocated to the off-ramps to ensure that, for all i, the queue on ramp i clears T_i time units before the queue on ramp 1.

Technology: Technology would have to play an essential role with dynamic off-ramp management because one would have to: a) decide when to close and open each exit based on

current traffic conditions and the storage capacity of the off-ramps; b) post messages on VMS's indicating which off-ramps are available; and c) coordinate the control with the local street network.

Variants for the end of special events: Similar distribution strategies could be tried at the end of special events, by redirecting event traffic to less used freeway on-ramps. However, this is not as promising. First, the benefits of this application would be smaller and would remain confined to stadium users. Second, this variant would be difficult to implement because: (a) it is challenging to inform everybody nearly simultaneously, and (b) drivers may not follow the suggestions once they are on the city streets.

4.3 Dynamic Merge Control

This strategy is based on untested elements of driver's psychology, so it is a speculative proposition that needs field-testing. It consists of controlling input flows in a merge bottleneck in order to maintain high discharge rates.

Preliminary empirical evidence shows that under certain conditions flows as high as 2,800 vphpl can be observed for long times on the median lane of a freeway, and that the total freeway flow can also be quite high (Cassidy and Bertini, 1999). However, as soon as a queue forms, a lower bottleneck capacity is observed. It appears that a precursor to queue formation is a slight drop in speed on the median lanes, followed by a slight reduction in discharge flow. If the entering vehicles cause the speed drop, then metering the on-ramp may restore high speeds, maintain high flows, and prevent the formation of a queue. If this approach (monitoring the median-lane speeds) does not work, one should try to identify other indicators that better correlate with the conditions that motivate people to drive so efficiently. The indicators can then be used to ensure that these conditions are maintained. To our knowledge, ways of sustaining high flows through metering have not yet been found.

Different theories offer different insights into the capacity-drop phenomenon. As mentioned in Zhang (2001), two theories that could explain it are Newell's asymmetric continuum theory (1962) and Daganzo's behavioral continuum theory (1999). These, and other theories, suggest key indicators that should be sensed and/or controlled. One should recognize, however, that indicators obtained from theories for long, homogeneous freeways may not apply when ramps are closely spaced. For example, if a busy downstream off-ramp is close to the bottleneck, the added friction

due to weaving may play a role. In cases such as this, one could try lane-assignment actions on the freeway (discussed in section 2) in conjunction with ramp metering.

Technology: Sensing speed, occupancy and flow, by lane, downstream and upstream of the merge would seem to be of critical importance, as is sensing the same measures on the on-ramp, but the specific items would depend on the indicators that turn out to be most useful.

4.4 Gridlock Management

This is perhaps the most beneficial application of ramp metering. The situation arises when a long queue spills back past many on-ramps and off-ramps, entrapping vehicles that wish to exit before the bottleneck; see Figure 6. If, in this case, one meters some of the on-ramps inside the queue, the bottleneck flow should not change. (The extra flow would be taken up by the vehicles queued upstream of the metered on-ramps.) However, with higher flows in the queue, more vehicles can exit and the flow on the upstream off-ramps would increase. Thus, the total system flow increases and conditions would improve. Note that this happens, even if the metered drivers have no alternative paths. [We call this the "gridlock effect" because it can lead to complete stoppages on closed-loop roads; a review of the gridlock effect can be found in Daganzo (1995). The effect also occurs on roundabouts, and this is why one should give priority to entering traffic on these facilities.]

Technology: Technology is needed here to: a) determine appropriate metering rates based on traffic conditions *upstream* of the metered location, and b) monitor the results taking into account the appropriate time lags and wave speeds.

5 MISCELLANEOUS CONTROL ACTIONS

5.1 Dynamic Speed Limits

Consider a two-lane freeway section with such high demand that a queue forms at its upstream entrance (the freeway is flowing at capacity). If the queue contains more fast vehicles than slow vehicles, and fast vehicles avoid the right lane because of the slow vehicles traveling in it, then

the right lane may carry little flow and the left lane on the order of 2200 vehicles/hr.[2] The queue would then be discharging at well under 4000 vehicles/hr. If a speed limit equal to the speed on the slow lane is imposed, fast vehicles will no longer avoid it. Although the flow on the left lane would be reduced (because of the lower speed) more fast vehicles would be observed on the right lane, increasing its density and flow. The combined flow (and the queue discharge rate) could increase, and eliminate the queue. If the freeway section is short, the total system delay could go down despite the increased travel time on the freeway section.

A research question is identifying flow compositions in the above-mentioned scenario where the benefits from an even distribution of flows and speeds across lanes easily outweigh the disbenefits from the reduced speeds on the fast lane(s). Another challenge is using this strategy on highways with congested entrances, where the design speed of the highway varies significantly due to changes in road geometry (e.g. long straight segments mixed in with hills and sharp curves). In this case speed limits might be desirable only in certain sections.

Other uses: Dynamic speed limits may also be useful in situations where freeways should be metered; e.g., major freeways that join a ring road around a city and can cause "gridlock" if allowed to inject too much flow into the ring. By imposing a temporary speed limit over a *short* freeway section, a freeway queue with a desired output flow can be created; i.e., the freeway is effectively metered. The section should be long enough to be able to bring the vehicles to the desired speed gently, but not any longer. Otherwise, the length of the bottleneck would be extended unnecessarily. [The benefit of "holding" vehicles has been recently recognized in the dynamic traffic assignment literature; Ziliaskopoulos (2000).]

Technology: Technology is needed to detect the traffic conditions that require the control (e.g., an inconvenient speed difference and low highway flows, in the first case) and to decide the 'best' control speed, based on current conditions.

---

2 Sites where this may be happening are: Westbound Interstate-80 from Lake Tahoe to Auburn on winter Sundays, and Southbound California State Highway 17, leaving Los Gatos, during the afternoon rush hour.

5.2 Rationing Free Access

Advanced technologies can also be used to reduce the demand for a congested facility, and not just to control traffic as previously discussed. Traditional approaches for reducing demand can be classified into two main groups: pricing, and coercive turn-taking schemes (e.g. based on license plates). These strategies are easy to implement, but they do not distribute their benefits and penalties equitably among the (potential) users. Therefore, they are not widespread. Fortunately, there is a scheme that overcomes this drawback, although it is not as easy to implement.

A hybrid of pricing and turn-taking in which people are allowed *paid* access all the time but are forced to take turns for *free* access has been recently shown to distribute its burdens and benefits more fairly (Daganzo, 1995, Garcia, 1999, Daganzo and Garcia, 2000 and Daganzo, 2000). This strategy can be used on major bottlenecks such as the Bay Bridge in San Francisco (California), and also to improve the utilization of HOV lanes in ways where everyone wins.[3] In this case, monitoring technology is essential to keep track of individual vehicles and update their accounts automatically. More enforcement details can be found in Daganzo and Garcia (2000).

5.3 Diversion Actions

Diversion actions are control strategies intended to prevent traffic from passing through a bottleneck by proposing alternative routes to specific subsets of freeway users. Three ways of diverting traffic away from a bottleneck, aimed at different user subsets, are envisioned:

5.3.1 Diversion of upstream freeway traffic

To reduce freeway queuing delays, it may be beneficial to post messages advising freeway drivers to take exits upstream of a bottleneck. The message should be addressed to those drivers who would stand to gain the most from being diverted; i.e., drivers about to join the queue, or recently in the queue, whose destination is close to the bottleneck. The information will be most useful to drivers who cannot estimate the travel times in the queue and on the alternative routes. Thus, it is likely to be most useful with non-recurring bottlenecks since in this case drivers cannot easily forecast their

---

[3] Since an obvious travel alternative exists, it can be shown that simple coercive rationing is also "Pareto efficient" for the HOV problem, and more equitable than the so-called "HOT" lanes.

queuing delay. The strategy should also be somewhat useful with recurring bottlenecks, since not every driver can be assumed to have prior experience.

A serious problem with this type of action is that one cannot easily anticipate how many drivers will act on the information. If too many do, new, more serious bottlenecks could be created at the off-ramps. Too much exiting traffic might also starve the original bottleneck for flow, which is also undesirable. Instead of fine-tuning the advice based on unreliable forecasts of "acceptance" rates, one should try to use "feed-back" information from the field to modify the advice. The acceptance rate can then be controlled by sending messages to a smaller or larger subset of drivers, and by turning the message on and off over time based on the observed response. [This strategy is very similar to the diversion measures of section 4.2, with the difference that now the congested off-ramp is not closed. The guidelines in Laval and Muñoz (2001) can again be used to set the diversion rates that would completely remove the freeway queue.]

5.3.2 *Diversion of entering traffic*

In the same spirit, it can also be advantageous to redirect traffic that would enter a congested freeway section directly upstream of a bottleneck to on-ramps downstream of the bottleneck. This can be achieved by metering severely, or closing altogether, those on-ramps that are directly upstream of the bottleneck, while at the same time advertising the delay at the entrance to the ramp and redirecting the affected traffic to the appropriate downstream ramps. This action can have a negative impact on surface streets, and therefore should be carefully evaluated beforehand. The impact can be mitigated with careful coordination of the freeway and local street controls. The strategy is likely to be most successful where the local street network is under-saturated and well connected.

With both diversion methods (Secs. 5.3.1 and 5.3.2) only people who can bypass the bottleneck at low cost should be asked to do so.[4] Therefore, VMS's could also specify for which destinations the diversion is recommended.

---

[4] Although it may appear that in the case of Sec. 5.3.2 all the entering traffic could be routed to downstream off-ramps, there are situations where the geometry of the system would imply a large penalty for some drivers if this were to be done. This would happen, for example, if there is a connection with another freeway immediately downstream of the bottleneck.

5.3.3   Diversion by pricing

A variation on these ideas uses money instead of time as the regulating incentive for the use of recurrent bottlenecks. It consists in setting tolls that depend on the origin and destination ramps of each freeway trip, in such a way that people who use the bottleneck are those for whom the travel time savings are greatest. Although there are equity issues, the availability of alternative routes should make this policy rather benign. The guidelines could be as follows:

1. Long-distance bottleneck users do not pay; i.e., when both the trip's origin ramp and destination ramp are far away from the bottleneck.

2. Short-distance bottleneck users pay; i.e., when either the trip's origin or destination is in the vicinity of the bottleneck. The toll could depend on proximity: the closer, the higher.

The modeling challenge here is to determine the appropriate (time-dependent) toll that would make the system to behave as desired. Because users are different and they might make different route choices on the surface streets, an accurate evaluation of benefits with detailed dynamic assignment models is not recommended. Nonetheless, if one can use simple models (without route choice) for the ramp-to-ramp trip times on the local street network, guidelines can be developed; e.g., along the lines suggested in Laval and Munoz (2001) for the case where drivers are identical. [This reference shows that if there is no congestion on the local streets, and one wishes to eliminate the freeway queues completely, the toll should increase very rapidly with time but decrease gradually.]

Technology: In all cases dealing with diversion (Sec. 5.3) technology is needed to detect the queue, estimate travel times, activate the signs properly, apply the tolls and monitor the results. Note that in all cases the traffic signals on the local streets should be re-timed to mitigate the effect of the diverted traffic. This requires coordinated control.

5.4   Dynamic Use of Shoulder Lanes

It is proposed here that shoulder lanes along *short* recurring bottlenecks should be used to increase their capacity. Shoulder lanes are already used in work zones and around incidents, but it is suggested here that they can also be used with recurring bottlenecks. This should only be done if shoulder design allows it. To be safe, a reasonable speed limit should be imposed through the bottleneck. The critical thing here is not that speeds should be high, but that flows across all lanes,

including the shoulders, should be higher than under the status quo, *despite the lower speeds*. A major difference with the work-zone application is that shoulder lanes should only be opened for use when needed; i.e., when the flow through the section exceeds the capacity of the no-shoulder-bottleneck, and whenever there is a queue. This could be determined in real-time.

Technology: Technology is obviously needed to detect the times when the shoulder lanes should be opened, when they are no longer needed, and to post appropriate messages and variable speed limits through the bottleneck.

5.5   Anonymous Automatic Vehicle Identification (AVI)

For many applications (both in traffic engineering and traffic science) it would be useful to know the space-time trajectories of all vehicles, without intruding into the privacy of their occupants. This information would enable us to accurately estimate the OD pattern on problematic freeway segments, and improve the control schemes.

Video sensors and machine vision algorithms, such as Liddy (2001), are a step in this direction. They can already produce reasonably accurate time-space vehicle trajectories from the field of vision of a single camera, and can also match vehicle features from camera to camera to extend the surveyed section indefinitely. Therefore, it will soon be possible to follow all the vehicles along a freeway instrumented with cameras. Although this should be useful for control purposes, the reliability of these systems is not yet high enough to issue fines for speeding and/or making illegal lane-changes.[5]

Some of the traffic control methods suggested above (e.g., in Secs. 2, 5.2 and 5.3.3) involve fines and/or the exchange of money. Therefore a more reliable AVI method would be helpful. In some cases, (e.g., Sec. 2 and 5.3.3) one only needs to know that a vehicle passing through two check points is the same, and a full vehicle identification is not needed. This is what we mean by anonymous AVI systems. The required information could be collected in active and passive ways:

Active: Vehicles could be equipped with identification chips that would contain a code readable by sensors, but the code would change every day. The chip could also contain an encrypted

---

[5] Video systems should first be deployed at locations where knowledge of the O-D table would result in the greatest benefits; ring-roads and problematic weaves are suggested here.

fixed code that could only be read by sensors authorized by the user, e.g., for the purposes of paying user fees. (An interesting piece of research would be to determine whether individual vehicles leave enough of a signature on the system, for their identities to be revealed after many repeated trips. If the answer is in the negative, perhaps the public could be convinced that this form of anonymous AVI carries a potential benefit.)

Passive: Perhaps a way can be found to "paint" or "tag" vehicles *temporarily*, as they enter a freeway, so that their paths can be tracked. This technology would obviate the need for tollbooths in turnpikes where the charges depend on the points of entry and egress, and would be very useful for some of the advanced control methods mentioned above.

6 CONCLUSION

This report has presented twelve strategies for improving traffic flow that require technology. Some are less conventional than others, but they all have in common that their success is easily measured, and it *does not depend on the validity of a large scale-model with innumerable assumptions and data needs*. The strategies are relatively simple and testable, but require the judicious use of technology. Severe field operational tests (FOT's), with clear answers, can be readily conducted.

Before this happens, however, a careful study of a candidate strategy should be conducted to achieve the following: (a) identify possible test sites and potential benefits; (b) establish viability from a human factors perspective; (c) find appropriate placement of sensors based on sound traffic engineering principles (recognizing wave speeds, time lags, etc…); (d) determine the optimum placement of signs and relevant messages; and (e) create the simplest possible control algorithms that will do the job. Since none of these tasks is technically daunting, and none of the strategies is prohibitively expensive, we hope that a fair number of them will soon be tested.

7 REFERENCES

Cassidy, M. and Bertini, R. (1999) "Observations at a freeway bottleneck." *Transportation and Traffic Theory*, Proc. 14th Int. Symp. On Transportation and Traffic Theory, pp.107-146 (A. Ceder, editor), Pergamon, New York, N.Y.

Daganzo, C.F. (1995): "A pareto optimum congestion reduction scheme" *Trans. Res. B*, 29 (2), 139-154.

Daganzo, C.F. (1995) "The nature of freeway gridlock and how to prevent it". Institute of Transportation Studies, Research Report UCB-ITS-RR-95-1; abridged in *Transportation and Traffic Theory*, pp 629-646, J.B. Lesort, editor, Pergamon-Elsevier, New York, N.Y.

Daganzo, C.F., Lin, W.H., Del Castillo, J.M. (1997) "A simple physical principle for the simulation of freeways with special lanes and priority vehicles". *Trans. Res. B*, 31 (2), 105-125

Daganzo, C.F. (1999) "A Behavioral Theory of Multi-Lane Traffic Flow Part I: Long Homogeneous Freeway Sections", Institute of Transportation studies Working Paper, UCB-ITS-RR-99-5. (http://www.ce.berkeley.edu/~daganzo/publications.html)

Daganzo, C.F. (2000) "Taking turns: Rx for Congestion" *Access*, Fall 2000, pp 14-19. (http://www.ce.berkeley.edu/~daganzo/publications.html)

Daganzo, C.F., R. Garcia, (2000): "A Pareto improving strategy for the time-dependent morning commute problem", *Trans. Sci.*, 34(3).

Erera et al (2001): "The access control problem on capacitated FIFO networks with unique O-D paths is hard", *Opns. Res.* (in press)

Garcia, R. (1999): "A Pareto improving strategy for the time-dependent morning commute problem". Ph.D. dissertation, Dept. of Civil and Environmental Engineering, University of California, Berkeley, CA, December 1999.

Laval and Munoz (2001): "System Optimum Diversion in Congested Freeways". Working paper (in progress).

Liddy, D. (2001): Private communication (work in progress).

Lovell, D. and C.F. Daganzo (2000): "Access control on networks with unique O-D paths", *Trans. Res. B*, 34(3) 185-202.

Muñoz, J.C. and Daganzo, C.F. (2000) "Experimental characterization of multi-lane freeway traffic upstream of an off-ramp bottleneck" PATH Working Paper 2000-13, July, 2000. (http://www.ce.berkeley.edu/~daganzo/publications.html)

Newell, G.F. (1962): "Theories for instability in dense highway traffic" *J. Opn. Res. Soc. Japan* 5, 9-54.

Ziliaskopoulos, A. (2000): "A Linear Programming Model for the Single Destination System Optimum Dynamic Traffic Assignment Problem," *Trans. Sci.* 34, 37-49

Zhang, H.M. (2001) "New perspective on Continuum Traffic Models". *Net. Spat. Econ.* 1, 9-34.
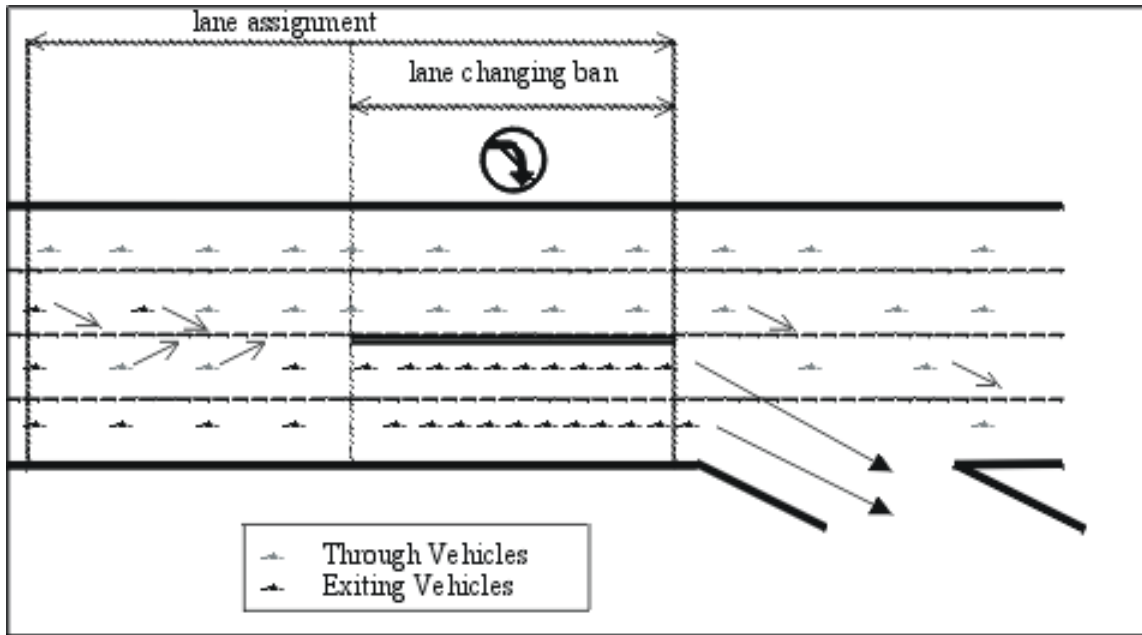
FIGURES

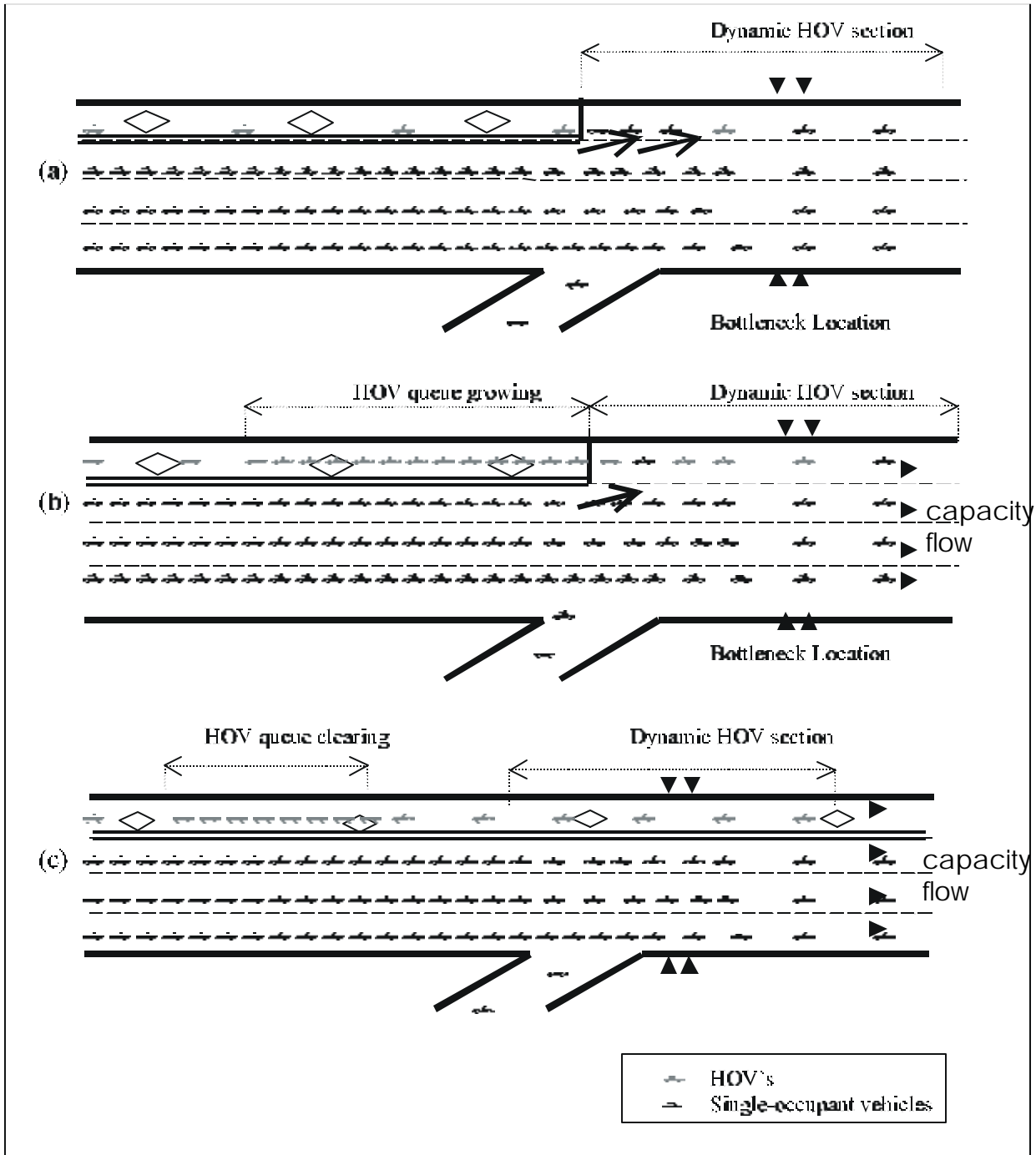**Figure 1:** Dynamic lane assignment plan upstream of a congested exit ramp.

Figure 2: A bottleneck with a dynamic HOV lane. Diamonds indicate when and where the HOV designation is in force. (a) Dynamic HOV designation just turned off (queue about to form); (b) HOV designation inactive (queue growing); (c) HOV designation active (queue clearing).
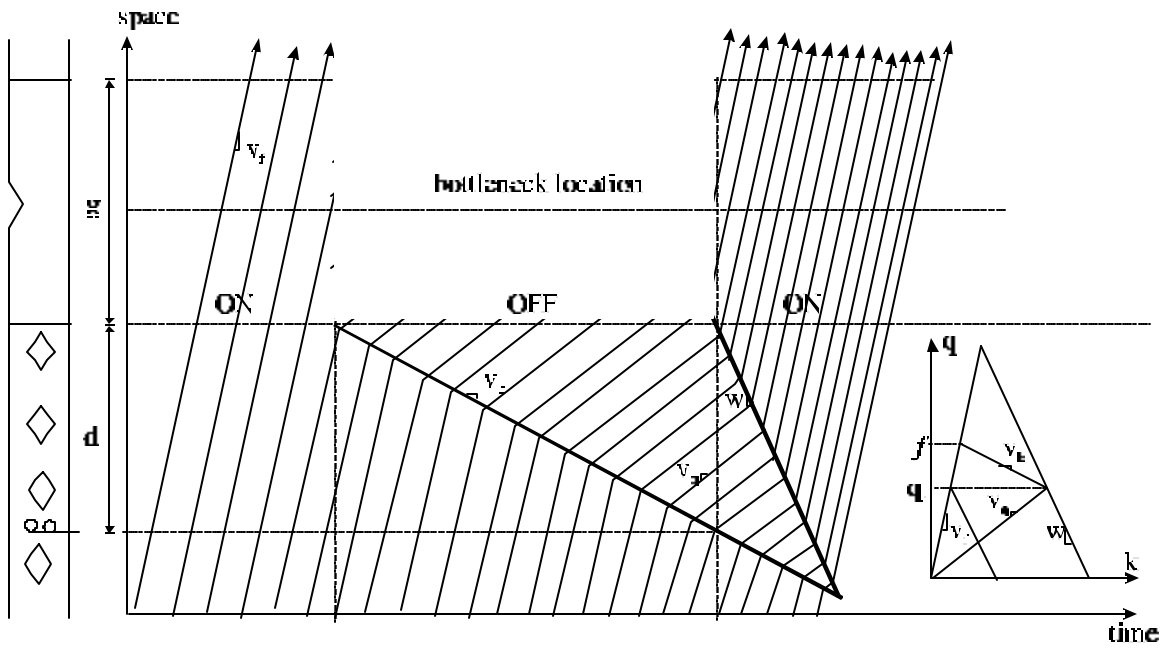
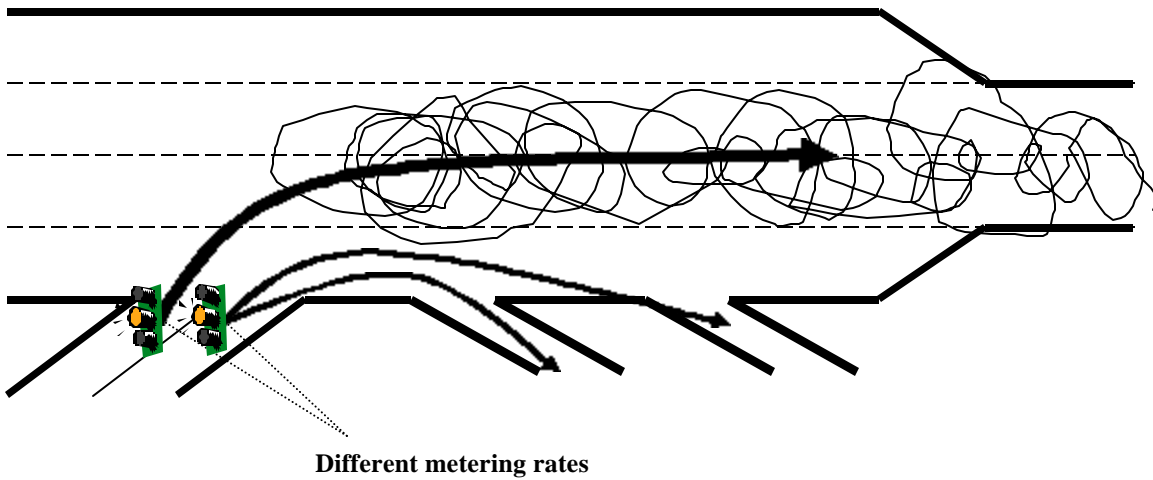**Figure 3: Space-time diagram for the HOV problem**

Different metering rates

Figure 4: Ramp-metering Actions / Destination specific metering at on-ramps
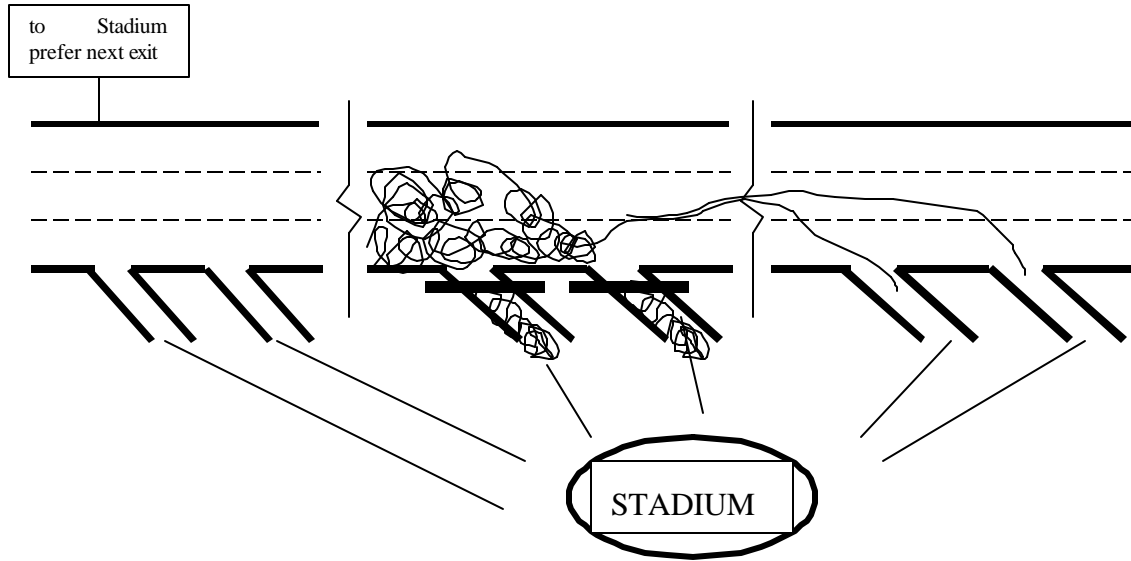
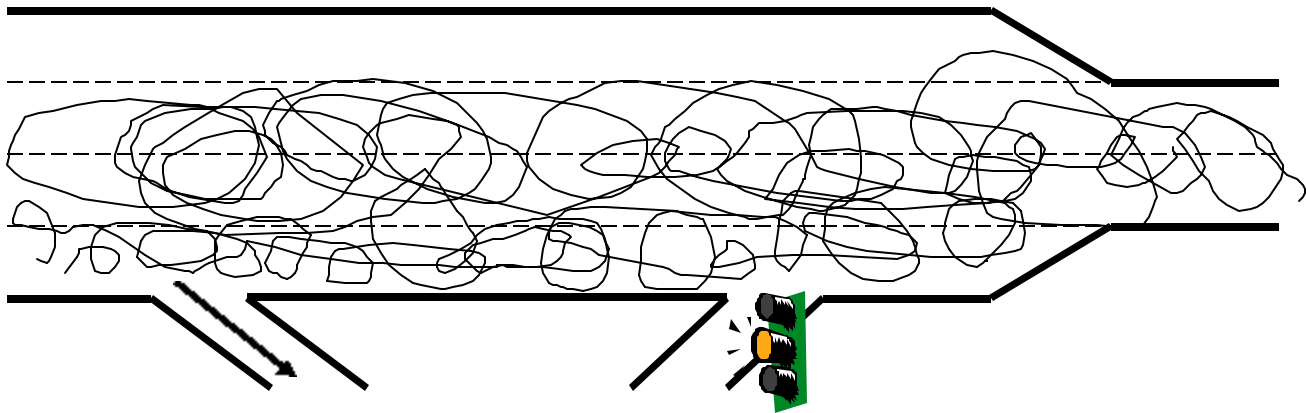to   Stadium
prefer next exit

STADIUM

Figure 5 : 4.2 Dynamic Off-Ramp Management

Figure 6 : 4.4 Gridlock Management