

QUEUE SPILLOVERS IN TRANSPORTATION NETWORKS WITH A ROUTE CHOICE

Carlos F. Daganzo
Department of Civil and Environmental Engineering
and
Institute of Transportation Studies
University of California, Berkeley, CA 94720

(January 18, 1996; revised, February 5, 1997, and August 14, 1997)
Transportation Science 32(1), 3-11, 1998.

Abstract

This paper explores some of the traffic phenomena that arise when drivers have to navigate a network in which queues back up past diverge intersections. If a diverge provides two alternative routes to the same destination and the shorter route has a bottleneck that generates a queue, one would expect that queue to stabilize at an equilibrium level where the travel time on both routes is roughly equal. If the capacity of the alternative route is unlimited then this network can accommodate any demand level.

However, if the bottleneck is so close to the upstream end of the link that the equilibrium queue cannot be contained in the link, then the trip time on the queued route cannot grow to match that on the alternate route. This means that the alternative route can never be attractive, even if the queue spills back past the diverge, and that drivers approaching the diverge will act as if the alternative route did not exist. As a result, a steady flow into the system greater than the capacity of the bottleneck will cause a queue to grow all the way back to the origin (blocking it). The final result is an "oversaturated static state" where the queue regulates the input flow into the system.

Curiously, if the bottleneck capacity of this network is reduced below a critical level (or is eliminated altogether) then the alternative route becomes attractive again and the system cannot reach the saturation point. This phenomenon is explored in the paper, where it is also shown that:

- (i) a network can become *permanently* oversaturated/undersaturated as a result of a *temporary* increase/decrease in link capacity,
- (ii) even under the most favorable assumptions, and in contrast to the equivalent assignment problem with point queues, a network can be stable both in an oversaturated and an undersaturated state, and
- (iii) temporary endogenous disturbances can permanently reverse the saturation state of a network.

These findings suggest that in certain situations the time-dependent traffic assignment problem with physical queues is chaotic in nature and that (as in weather forecasting) it may be impossible to obtain input data with the required accuracy to make reliable predictions of cumulative output flows.

INTRODUCTION

This paper is concerned with the distribution of traffic over congested networks recognizing the presence of physical queues, i.e., queues that take up space. Both steady state, i.e., static or time-independent, and dynamic issues will be discussed. A goal of this paper is to illustrate that the effects of queue spillovers past diverge junctions change the nature of the traffic assignment problem in a substantive way.

[The terms “traffic assignment” (TA) and “dynamic traffic assignment” (DTA) are used in this paper to describe the time-dependent evolution of a network in which users behave *non-cooperatively* according to certain rules, and where the cumulative input flows as a function of time are given for every origin/destination (O/D) pair. The solution of the DTA problem can be characterized by a set of cumulative O/D flow curves upstream and downstream of every link (called here the N-curves) and at the final destinations (called here the output curves). The term “static traffic assignment” (STA) is used for the case where the input and output O/D flows are time-independent. These TA scenarios encompass networks with traffic responsive control schemes, but do not apply to imaginary systems where users would cooperate for the benefit of the whole. The words “cooperative traffic assignment” or “system-optimum traffic assignment” will be used to refer to these scenarios when necessary.]

It is already known that spillovers past merges can lead to gridlock on ring roads and other networks with closed loops and that, as a result, the time-dependent output flows (by O/D pair) for networks susceptible to gridlock may be very sensitive to the time-dependent input flows; e.g., where a slight increase in the feasible O/D flows of a gridlock-susceptible road can result in no output flow at all (Daganzo, 1996a). The possibility of lockup also means that the DTA solution may not exist, even with a finite population of users. This is in contrast to the problem with point queues, where the DTA solution is known to exist under fairly mild conditions (Smith, 1993). It is also known that gridlock can be triggered by expanding on-ramp capacities, and prevented by metering them.

This paper will show that similar, and in some respects more problematic, changes in the character of the traffic assignment problem arise when queues back up past diverges. The paper

is organized as follows. Section 1 discusses the DTA problem for a network with one O/D pair and two parallel links. The section shows that an increased capacity for one of the links can lead to a reduced output and to oversaturation; i.e., to the non-existence of a static equilibrium assignment in the conventional sense. The prevalence of this and similar phenomena is also discussed in this section. Section 2 then shows that both oversaturated and undersaturated static states are possible on the same static network. Section 3 examines the system's evolution toward these states from a set of initial conditions, as well as the nature of temporary perturbations, both exogenous and endogenous, that can change the saturation state of the static network. The implications that these findings have for time-dependent problems and traffic research are also discussed.

1. PHYSICAL QUEUES

The phenomena of interest arises when drivers approaching a diverge can reach their destination using either one of the branches. It will be shown in this section that if the capacity of one of two links in parallel is increased, its queue may grow longer. This is true whether one neglects the spatial dimension of the queue or not. It will also be shown that the spatial extent of the queue may cause the network to become oversaturated, so that by increasing the capacity of a link the "network capacity" is reduced. This will be done in two ways. First by showing in detail how a final static state is reached over time from a set of initial conditions and then (more easily, but less intuitively) by looking at the time-independent conditions that must be satisfied by an undersaturated equilibrium, if one exists.

1.1 Transition toward a final static state

Consider the network of Fig. 1, where travelers from "O" to "D" can choose either link, $i = 1$ or 2 . Link 2 is shorter in time than link 1 but includes a bottleneck of capacity, c_2 . The time needed to traverse link i when there is no queue, denoted τ_i , is assumed to be independent of flow. It is also assumed that on arriving at node "N", drivers evaluate accurately the added queuing time that they will experience on link 2, e.g., based on what they can see ahead, and then choose the link that will get them to "D" earliest. This is the Wardropian model of driver

behavior (Wardrop, 1952) which has been used extensively in the DTA literature; see for example Kuwahara and Akamatsu (1993), Smith (1993), and Heydecker and Addison (1996).

[Wardrop's model requires that drivers anticipate the future perfectly for the duration of their trips. This is not unreasonable for a small network such as ours but may not be very realistic for large congested networks involving long trips. On the other hand, models without anticipation lead to unreasonable results and cannot be considered as serious alternatives; see Daganzo and Lin (1995). Since a clearly superior model of route choice does not exist, Wardrop's model is adopted in this paper despite its shortcomings. It should be stressed, however, that the qualitative conclusions of this paper also follow from other behavioral models, e.g., models with imperfect information, lazy drivers, heterogeneous drivers, etc...]

The evaluation of queues, delays, and cumulative output for a given time-dependent set of arrivals on the network of Fig. 1 is a standard problem in queuing theory; see problem 3.4 in Newell (1982). Figure 2 displays the solution to this problem in terms of the standard N-curves of cumulative count, for the case where a steady flow, $Q > c_2$, is suddenly released into an empty system. The figure starts at the time when the first vehicle arrives at "N". The top half displays the N-curves observed at the upstream and downstream ends of link 2, labeled U_2 and D_2 , and the bottom half the corresponding curves for link 1. The curves labeled U_T and D_T on the top part are the sum of both link counts, i.e., the input and output O/D curves. They represent the total cumulative count observed upstream of node "N" and at "D".

Note that the trip times on link 2 of the vehicles arriving prior to time t' increase with successive vehicles. Of course, if that time were to exceed the trip time on the alternative link, τ_1 , by a significant amount most drivers would opt for the alternative route. Here we assume that drivers are very sensitive to time differences and that the diversion occurs as soon as τ_1 is exceeded by an infinitesimal amount. This occurs at point "P" of the figure (arrival time t'). From then on, the drivers individual guesses and choices of the fastest route should have the effect of roughly keeping the same trip time on both routes. In our Wardropian simplification the trip times should be precisely equal, and therefore take the value τ_1 on both routes. Since curve D_2 must have a constant slope as long as the bottleneck is at capacity, curve U_2 has to bend at point "P" and become parallel to D_2 . The flow so diverted then materializes on link 1; see bottom of figure.

The figure shows that eventually, after the time corresponding to point “S”, the system reaches a steady state where the three pairs of input-output lines become parallel.

The vertical separation between curves U_2 and D_2 represents the number of vehicles in link 2. It stabilizes at a value, n_2 , that is assumed not to exceed the maximum storage capacity of the link. The magnitude of the latter, denoted \hat{n}_2 , is depicted by a vertical double arrow in the top part of the figure.

The storage capacity depends on the bottleneck service rate and the location of the bottleneck within the link. We know from traffic flow observation that if a bottleneck is made more restrictive, then the density of queued traffic upstream of the bottleneck increases and the density of traffic downstream decreases. It should thus be clear that \hat{n}_2 can change if c_2 is modified. However, the direction of the change depends on the position of the bottleneck within the link. In view of this, \hat{n}_2 is assumed to be independent of c_2 in our example; however, the phenomena about to be illustrated also arises if it varies.

If in an effort to improve the system an analyst now decides to improve the service rate of the bottleneck, it should be easy to see from the construction of Fig. 2 that the equilibrium would be reached later and that the value of n_2 , which is $n_2 = \tau_1 c_2$, would then go... up! A longer queue would be observed as a result. This is not all, however. Too much of an improvement can lead to more serious problems, as is explained below.

If in our example c_2 is increased by 33% then the result of Fig. 3, which has been drawn to scale, is obtained. The figure shows that the occupancy of link 2 reaches its maximum value \hat{n}_2 before the trip times have grown enough to match those of the alternative route. At that time, when the queue has grown all the way to node "N", the trip time on route 2 is $t_2 < \tau_1$ and the alternative route is still unattractive. If no one chooses it, then only as many vehicles will flow past node "N" as are allowed to enter link 2. Therefore, the total flow at "N" will be c_2 and a queue will grow on the approach to that node. This is indicated by the diverging curves of desired (or virtual) arrivals at "N", V_T , and the actual N-curve, U_T . We see, thus, that the desired demand could not be carried by this network. Over time, the queue on the approach to the diverge would grow all the way to the origin, blocking it and reducing the input flow to c_2 ; the network would then have reached a static state where the throughput is less than the demand.

As in the case of gridlock, this is an instance where an improvement in an undersaturated network leads to oversaturation. Although the effects of queue spillovers past diverges are less drastic than the gridlock effect (output flows in the present case remain positive), they are more difficult to control by doing something at the junction because people cannot be easily induced to choose a long route when a better choice is available. In our example it is easy to imagine a real (non-Wardropian) driver arriving at the diverge after spending a long time in the approach queue and thinking “I can now take a long detour, $t_1 = \tau_1$, or else just wait a little longer in the queue and be done with my trip in $t_2 = \hat{n}_2 / c_2$ ”. Even if the driver recognizes that taking the longer route may be better for those behind, having already paid his/her dues in the queue, (s)he would probably decide to push through the bottleneck anyway; especially, if the difference $t_1 - (\hat{n}_2 / c_2)$ was significant. This illustrates that perfect Wardropian behavior is not necessary to generate the phenomena discussed in this paper.

For Wardropian drivers, consideration of Figs. 2 and 3 reveals that if the bottleneck capacity is greater than a critical value, $c_{2,crit} = \hat{n}_2 / \tau_1$, then the situation of Fig. 3 arises and the system becomes oversaturated by any traffic inflow greater than the bottleneck capacity. On the other hand, if the bottleneck capacity is smaller than the critical value, then the situation of Fig. 2 arises and the system cannot reach saturation.

1.2. Conditions for an undersaturated equilibrium

The conclusion that expansion of a bottleneck can reduce capacity can also be reached without examining the time-dependent behavior of the network, simply by stating the conditions that must be satisfied by an undersaturated static equilibrium and then verifying its existence or lack thereof. (The word equilibrium is used in this paper, as in the STA literature, exclusively to denote undersaturated states.)

Let t_i denote the equilibrium trip time including queuing delay on link i , and q_i denote the equilibrium flow. The equilibrium conditions are then as follows. First, flows must be non-negative and satisfy the flow conservation and capacity constraints:

$$q_1 + q_2 = Q \quad \text{and} \quad q_2 \leq c_2 . \quad (1)$$

In addition, the trip times that can arise must be consistent with the queue development condition:

$$t_2 \geq \tau_2, \quad \text{with } t_2 > \tau_2 \text{ only if } q_2 = c_2 \quad (2)$$

$$t_1 = \tau_1. \quad (3)$$

And, according to Wardrop's principle,

$$t_i = \min(t_1, t_2) \quad \text{if } q_i > 0. \quad (4)$$

Equations (1)-(4) correspond to a model with point queues. They have a solution for every possible Q . In the case of interest here, $Q > c_2$, the unique solution of (1)-(4) is:

$$q_2 = c_2, \quad q_1 = Q - c_2, \quad t_1 = t_2 = \tau_1. \quad (5)$$

This equilibrium is consistent with Fig. 2.

If space is a consideration one must also check that the equilibrium accumulation, given by Little's formula of queuing theory, does not exceed the maximum possible, i.e., that:

$$n_2 = q_2 t_2 \leq \hat{n}_2. \quad (6)$$

This constraint is the difference between a static model with point queues and one with physical queues. When $Q > c_2$, q_2 and t_2 are given by (5), and (6) becomes:

$$c_2 \tau_1 \leq \hat{n}_2 \quad (7)$$

which is a relation between constants.

If inequality (7) is not satisfied, i.e., $c_2 > c_{2,\text{crit}} = \hat{n}_2 / \tau_1$, then a static undersaturated equilibrium cannot exist when $Q > c_2$. If the inequality is satisfied, however, then (5) and (6) define a valid equilibrium. These results are consistent with the findings of the previous subsection.

1.3. Prevalence of the effects

The above is an example of an assignment problem where a feasible solution that satisfies the flow conservation equations and the link capacity constraints (1) exists but the equilibrium does not. This undesirable effect also arises in the standard STA problem with point queues when the network junctions are controlled by certain traffic-responsive algorithms (Smith, 1979a). What is noteworthy now, is that spillovers generate the phenomenon on a very simple network with a very simple cost/capacity structure, and this suggests that the effect may be quite common.

Note as well that the system optimum assignment also exists for our simple network: one would simply send enough flow on link 2 to keep the bottleneck on the verge of oversaturation, without building a queue, and would send the rest on link 1. Thus, the example also illustrates that drivers' choices at diverges can be so inefficient from a collective standpoint so as to make an "infinite" difference in the total travel time.

[These phenomena cannot happen in networks with point-queues if, as in our case, the trip time on a link only depends on its own flow and this flow is limited by a fixed capacity. This is true because for models with point queues the feasible point can be the initial point for a converging equilibrium seeking algorithm; see Daganzo (1977), for example, and the more general results in Smith (1979b).]

It should also be noted that if in our example the link with the bottleneck is removed altogether, then an equilibrium becomes feasible. As such, the example of Fig. 1 extends the scope of the diversion "paradox" (Braess, 1968), in that not just a reduction in delay but the feasibility of an equilibrium can be obtained by eliminating a link in a very simple network with a route choice: one with just one O/D pair, one bottleneck and two parallel links.

The effects illustrated by the network of Fig. 1 happen in practice quite commonly; e.g., when some of the drivers on a freeway cannot be persuaded to use the local streets to get around a bottleneck and thus cause upstream queues, and also when an oversaturated narrow off-ramp, e.g., leading to a football stadium, generates a freeway queue. If in the latter case a much wider but less convenient ramp exists somewhere downstream, then undersaturation can be restored by closing the narrow off-ramp.

When a network includes more than one O/D pair, some O/D pairs can be undersaturated

and others not. Figure 4 is a schematic representation of a common situation where one can restore undersaturation for one O/D pair, even if it is not possible to restore it for all. Links 1 and 2 of that figure represent the two parallel lanes of a one-directional two-lane freeway going from “O” to “D₂” and link “c” represents an exit ramp leading to “D₁”. Links “a” and “b” are two of the many links that could be used to represent lane-changing. If the flow of exiting vehicles exceeds the capacity of link “c” on this network, then a queue of exiting vehicles will develop on the freeway. In a cooperative system-optimum assignment, this queue would be confined to the right lanes (link “2”) and the through traffic would proceed unhindered on lane “1”. However, if this was the state of affairs and drivers were then “decontrolled”, some exiting drivers would find it attractive to travel on link “1” and cut in the front of the queue (using link “b”). This would be the Wardropian behavior. If *some* drivers behaved in this way, they could create a queue on links “b” and “1” that would entrap the through vehicles. (Note that this would not happen in a model with point queues because the exiting queues would be confined to links “b” and “2”.) The spillover queue could reduce the maximum flow for the through trips below their demand as well, and the whole freeway queue could grow. We see thus that eliminating link “b”, e.g., by lane markings and enforcement, would restore undersaturation for the through trips.

These three examples illustrate that the blocking effects introduced by queuing increase both the magnitude of the “diversion effect” and the difference between the cooperative and non-cooperative static states that arise, and also that the effects are quite prevalent and relatively independent of how drivers choose routes.

The next two sections show that static networks with physical queues can be stable both in an undersaturated and a saturated state, and that temporary perturbations of the O/D flows can cause a shift from undersaturation to oversaturation and vice versa. This suggests that the cumulative output O/D flows in very congested time-dependent networks can be very sensitive to minor fluctuations in the cumulative input O/D table. Section 2 focuses on the final static states, and Sec. 3 on state transitions.

2. COEXISTENCE OF OVERSATURATION AND UNDERSATURATION

The equilibrium link flows of the static traffic assignment problem are unique if the link

time functions are monotone in the link flows; see Smith (1979b). This uniqueness property extends to networks with fixed link costs and point queues caused by link bottlenecks of fixed capacity, since the latter problem is a limiting case of the former. Yet, if physical queues are introduced into the point queue problem by means of storage constraints such as (6), then a multiplicity of feasible equilibria can arise. This is well known; see Newell (1977).

This section extends these results by showing that a network with physical queues can be stable both in a feasible and infeasible static state. A “static state” is defined as an O/D route flow pattern where no driver can find a better route. It can be either undersaturated (a feasible equilibrium) or oversaturated (an infeasible static state). We have already said that the latter case arises if a queue blocks one or more of the origins and prevents them from discharging the required flow. The distinction between the two kinds of static states is important because in one case the O/D outputs equal the desired inputs and in the other they do not.

The situation of interest arises in the network of Fig. 5A where Q_1 units of flow go from O_1 to "D". Although the figure also includes another origin, O_2 , it will be assumed for the time being that the second origin is inactive. By using only one O/D pair, the example will establish that infeasible static states can arise that are Pareto inefficient relative to the also feasible equilibria, in the sense that *all* network users are “infinitely” worse off in one static state than in the other.

A roundabout is used in Fig. 5A as a metaphor for other types of junctions in which the traffic from link 2 would have to yield the right of way to the traffic from link 1. Alternative scenarios are depicted in Figs. 5B and 5C. In general the situation arises when a longer, higher capacity route emanating from within a queue can be used to cut in front of it. The network of Fig. 4 is another example.

It is assumed that the bottleneck of link 2 is at its downstream end, and that its maximum service rate, c_2 , depends on the availability of traffic gaps in the roundabout’s priority stream, the flow of which is denoted $q^p = q_1$. Logically, there should be a non-increasing relation between c_2 and q^p ,

$$c_2 = c_2(q^p); \quad (8)$$

its particular form should depend on the type of junction. For stop signs and low speed merges (8) drops sharply first and more gradually later, as shown by the *capacity curve* of Fig. 6. For high speed merges without a clear priority the curve may start flat and then drop sharply with a slope close to -1; in this case the curve should only be used for low cross-flows, up to the point $q^{p,crit}$ where the cross-flow is also restricted by the interactions at the merge and mainline queues begin to develop.

It is now possible to see that the two static states (A and B) of Fig. 7 are possible. In that figure a dashed line represents a road segment with no flow, a double line a queue and a solid line unqueued flow. State A is the oversaturated case discussed in Sec. 1; we already know that it arises if:

$$Q_1 > c_2(0) > c_{2,crit} . \quad (9)$$

The growing queue on the approach to node N_1 would eventually block the origin and as a result its flow would be reduced.

Since state B has a queue that is confined to link 2, its flows must satisfy

$$q_2 = c_2(q_1) < c_{2,crit} . \quad (10a)$$

The equality of (10a), together with the condition

$$q_2 = Q_1 - q_1 , \quad (10b)$$

determines the equilibrium flows; they are the coordinates of the intersection point of the *demand* and *capacity* curves of Fig. 6. The inequality of (10a), which ensures that the equilibrium queue remains confined to link 2, is satisfied if the equilibrium point of Fig. 6 is below the displayed critical line.

Assuming that the capacity curve decreases as shown in Fig. 6, we see that two static states are possible if Q_1 is large enough. Section 3 shows how both static states can arise

naturally from a given set of initial conditions.

It should be stressed that the number of static states that can be reached in more complex networks should grow very rapidly (combinatorially) with the size of the network, as measured by the number of O/D pairs, bottlenecks and links. Evidence in this respect can be found in Newell (1977), which describes the many static equilibria that can arise in a simple network consisting of two parallel routes with crossover links. The new point being made here is that, in addition to many equilibria, a system may also be stable in a number of oversaturated states. This is important because the system outputs are equal to the desired inputs in the undersaturated case, but not in the oversaturated case.

A consequence of stability is that any incentives used to drive the system toward one static state from another only need to be temporary. This is unfortunate for the purposes of traffic forecasting because it means that spontaneous jumps from one state to another can also be caused by temporary flow disturbances. This is explained in the following section where it is shown that temporary changes in flow can cause the simple system of Fig. 5A to switch (either way) between undersaturation and oversaturation. To keep the presentation simple, as few details of the time-dependent solution during the transitions are presented as possible.

3. INDUCED AND SPONTANEOUS TRANSITIONS

Figure 3 showed that if the flow Q_1 was released on an initially empty network then the state of Fig. 7A would be reached because link 1 is always avoided. Using the same logic, we see that if link 2 was partially blocked during the transient phase (e.g., by flow from O_2) so that $c_2 < c_{2,crit}$, then traffic would be diverted into link 1, as in Fig. 2, and the state of Fig. 7B could be reached. Thus, both states can arise. It is shown below that the system can also jump from one state to the other.

For example, a transition from state 7A (oversaturation) into equilibrium 7B can be induced by temporarily restricting or interrupting the flow of link 2 to encourage the use of link 1 until pattern 7B is established. This result is interesting in that by making a network link *temporarily* worse, a *permanent* improvement in the network capacity is obtained.

Figure 8 illustrates the process by displaying two of the intermediate states that would arise

during the transition if the link is closed at its downstream end and drivers are informed of the consequences at the upstream end.

Clearly, the closure would force the vehicles at the head of the link 2 queue to stop and would encourage approaching drivers to divert to link 1; the diversion would occur at the maximum rate possible, which we denote Q_{\max} . [In a Wardropian world drivers would know beforehand when the closure was going to occur and would begin to divert sooner; no queue would be formed on link 2. In order to illustrate that Wardropian driver behavior is not necessary for the transition to occur, it is assumed here that the closure is not anticipated. The reader can verify that the transition also occurs in the Wardropian case.] Part A of the figure depicts the state of the system, shortly after the closure; it shows the clearing wave upstream of node N_1 , as well as the 3 shocks that mark the end of the two queues and the (darkened) portion of the road containing stopped vehicles. The dashed arrow directly downstream of N_1 is used as a reminder that there is no flow into link 2.

Eventually, all the vehicles of link 2 would come to a stop and the clearing wave upstream of N_1 would catch up with the back end of the upstream queue, or would reach the origin node, dissipating the queue. This event would restore the flow on the approach link to level Q_1 , and also on link 1. The resulting state of affairs is displayed on part B of the figure. Equilibrium pattern 7B should then arise from this state on removal of the restriction.

The reverse effect can also occur. If $c_2(0)$ is temporarily increased by some means, e.g., by using a police officer to replace the stop sign with a yield sign, then a jump from pattern 7B to 7A (oversaturation) is possible.

This undesirable outcome illustrates the danger and likely inefficiency of traffic control schemes that ignore route choice. Unfortunately, as is explained in the concluding subsection below, the time-dependent, link-by-link traffic distribution (by destination) in a network with spillovers and route choices may sometimes be very difficult if not impossible to predict. In these cases, improved traffic control schemes that recognize route choice but base the results on predictions of the destination- and time-specific traffic distribution, e.g., on a “rolling horizon”, cannot be expected to produce efficient solutions either; especially in congested cases, where the traffic distribution has to be predicted over a very long horizon.

Spontaneous transitions and sensitivity to the input data

Just like a temporary alteration to c_2 can induce a transition from undersaturation to oversaturation and vice versa, it is also possible for fluctuations in flow from O_2 to D (see Fig. 5A) to induce transitions in either direction.

To see this, imagine that there is some background flow from O_2 to D , Q_2 , and that the expression for $c_2(q_1)$ used at the outset of this section already incorporates the effects of said background flow. Then, fluctuations in Q_2 will induce temporary changes in the bottleneck capacity that can be of the type described above. Thus, surges in traffic from D_2 have the potential for inducing transitions from pattern 7A to 7B, and lulls in traffic for having the opposite effect-- a more detailed description of these transitions is given in a prior version of this paper (Daganzo, 1996b). The same logic reveals that flow fluctuations in the background traffic from O_2 can also influence the way traffic from O_1 is served by the network when the demand from O_1 changes with time, e.g., as in the example of Figs. 2 and 3.

It is beyond the scope of this paper to discuss all the possible types of network geometries (e.g., including multiple destinations) and flow fluctuations that could induce transitions, i.e., have large effects elsewhere, but it should be noted that the ones just presented are not exceptional. For example, transitions in one direction or the other can also be induced by fluctuations in the cross-flows of streets (not represented in our figures) that would intersect any of the links in our network, e.g., interfering with the queue downstream of node "N", and by fluctuations in the arrival flow itself. Note as well from the discussion of Fig. 8 that whether the transitions actually take place or not depends on the magnitude and duration of the traffic pulses and, more specifically, that: (1) a disturbance needs to have a minimum size to induce a transition, and (2) the smaller the network in terms of distances, trip times and storage capacities, the smaller the critical size. As a rule, transitions should be least likely to arise in networks of widely spaced intersections.

Since fluctuations in flow can have a lasting, and therefore large effect on the cumulative flows somewhere else in the network, and since large changes in cumulative flow are more likely to induce other transitions, we see that the effects of a fluctuation can trigger another transition

elsewhere. If the particular network geometry favors this chain-reaction then the overall system may be “unstable”. In view of this, one should not be surprised if in a particular DTA application with spillovers one finds that small changes to the time-dependent O/D inputs have a large impact on the O/D outputs and on the distribution of traffic on the network (by destination) at any given time.

The hypersensitivity of time-dependent, destination-specific link flows to input data should be more likely in a complex network with many O/D's, many bottlenecks, short links and interacting queues since there are more opportunities for the required conditions to arise and for widely different chain reactions of events to be set in motion from similar initial conditions and data. This should be expected because in a complex network the solution will depend on the relative arrival times of congestion pulses at the various intersections and on the order in which links spill over.

Whenever hypersensitivity arises, one should not expect to obtain very meaningful predictions and/or control strategies from even the best detailed models of dynamic traffic assignment. [The hypersensitivity issues raised here do not necessarily extend to the system-optimum DTA problem with physical queues; this problem can be shown to be a convex program for certain scenarios that include Fig. 1 as a special case; see (Ziliaskopoulos, 1997) and (Lin and Cao, 1997).] This author does not know what the answer is to these difficulties. It seems reasonable to posit that one should focus on searching for strategies that avoid spillovers. This, of course, is more easily said than done for freeways, but possible for surface streets. For example, if an inner city with closely spaced intersections experiences heavy flow and is susceptible to capacity-reducing spillovers, it may be wise to draw a line around the city center and restrict entering traffic to the levels that can be sustained by the network; in this way queues would be moved away from the city core and would do less damage. It is this author's opinion that the game in congested network control is queue management with limited information, and that any strategy that does not allocate space effectively is unlikely to succeed. If spillovers can be prevented, perhaps one should choose only among strategies that avoid them, since this leads to better flow and also allows the alternative control scenarios to be analyzed and fine-tuned with fewer difficulties.

Acknowledgment

Research supported by the U.S. and California Depts. of Transportation through PATH MOU-249.

REFERENCES

D. Braess, "Über ein paradox de verkehrsplanung", *Unternehmenstorchung* 12, 258-268 (1968).

C. F. Daganzo, "On the Traffic Assignment Problem with Flow Dependent Costs II," *Transportation Research* 11, 439-442 (1977).

C. F. Daganzo, "The nature of freeway gridlock and how to prevent it", in *Transportation and Traffic Theory*, pp. 629-646, J.B. Lesort, editor, Pergamon-Elsevier, New York, 1996(a).

C. F. Daganzo, "Effect of queue spillovers on transportation networks without a route choice", Institute of Transportation Studies, Research Report, UC-ITS-RR-96-1, Univ. of California, Berkeley, 1996(b).

C. F. Daganzo, and W. H. Lin, "The effect of modeling assumptions on the evolution of queues in a single corridor", *Transportation Research Record* 1453, 66-74 (1995).

B. G. Heydecker and J. D. Addison, "An exact expression of dynamic equilibrium" in *Transportation and Traffic Theory* (J.B. Lesort, editor) pp. 359-384, Pergamon-Elsevier, New York, 1996.

M. Kuwahara and T. Akamatsu, "Dynamic equilibrium assignment with queues for a one-to-many OD pattern" in *Transportation and Traffic Theory* (C.F. Daganzo, editor) pp.185-204, Elsevier, New York, 1993.

W. H. Lin and J. M. Cao, "A system optimal dynamic traffic assignment model with limited system control", *Transportation Research Board* presentation preprint, Washington, D.C., 1997.

G. F. Newell, "The effect of queues on the traffic assignment to freeways" *Proceedings 7th International Symposium on Transportation and Traffic Theory* (T. Sasaki and T. Yamaoka, editors) pp. 311-340, Kyoto, Japan, 1977.

G. F. Newell, *Applications of queuing theory* (2nd ed.), Chapman Hall, London, 1982.

M. J. Smith, "Traffic control and route choice; a simple example" *Transportation Research* 13B, 289-295, 1979(a).

M. J. Smith, "The existence uniqueness and stability of traffic equilibria" *Transportation Research* 13B, 295-304, 1979(b).

M. J. Smith, "A new dynamic traffic model and the existence and calculation of dynamic user equilibrium on congested capacity-constrained road networks", *Transportation Research* 27B, 49-64, 1993.

J. G. Wardrop, "Some theoretical aspects of road traffic research", *Proceedings Institute of Civil Engineers, Part II*, 1, 325-362; Discussion, 362-378, 1952.

A. K. Ziliaskopoulos, "A linear programming model for the single destination system optimum dynamic traffic assignment problem", *Transportation Research Board* presentation preprint, Washington, D.C., 1997.

LIST OF FIGURES

1. The simplest possible network with a route choice.
2. Development of an equilibrium on a simple network.
3. Development of an oversaturated state when the bottleneck capacity is 33% larger.
4. Schematic representation of a two-lane freeway and an offramp.
5. Three scenarios susceptible to multiple static states: (A) bypassing a roundabout queue; (B) bypassing a congested left turn around a city block; and (C) bypassing a congested freeway lane drop.
6. Conditions for the existence of an undersaturated equilibrium and an oversaturated static state.
7. Two static states that can arise from the same demand on the simplest network, if it terminates near a priority junction: (A) oversaturation; (B) undersaturated equilibrium.
8. Intermediate states encountered in a transition from oversaturation to undersaturation that is achieved by a temporary closure of link 2.















