UC Berkeley
Research Reports

Title

A Network Model of Departure Time Choice with Spillovers and Merging Effects. Part I: Building Block

Permalink

https://escholarship.org/uc/item/1rw9v116

Authors

Lago, Alejandro
Daganzo, Carlos F.

Publication Date

2003-07-01

Institute of Transportation Studies
University of California at Berkeley

A Network Model of Departure Time Choice with Spillovers and Merging Effects. Part I: Building Block

Alejandro Lago and Carlos F. Daganzo

A Network Model of Departure Time Choice

with

Spillovers and Merging Effects.

Part I:Building Block

Alejandro Lago and Carlos F. Daganzo

Dept. of Civil & Environmental Engineering,
University of California at Berkeley, Berkeley, California 94720-1720.

October 2002

Abstract

This paper presents a departure-time user equilibrium model that explicitly considers the most important determinants of congestion behavior in cities during the morning commute: different commuter origins, merge interactions and queue spillovers. The proposed model combines three previous works: the departure-time equilibrium theory in Vickrey (1969), the traffic flow model of Newell (1993) and the merge theory in Daganzo (1996). The paper examines the simplest possible network exhibiting the three important features and discusses the ensuing policy implications. The solution algorithm can be used as a building block for equilibrium analysis of more complex, single-destination networks with departure-time choice. The results reveal unexpected situations where ramp-metering can be beneficial, and others where the provision of more freeway storage can be counterproductive.

Introduction

Vickrey (1969) describes the first traffic model where commuters can adapt their departure time to avoid periods of high congestion. The model is very simple – a single bottleneck with a fixed number of commuters – but also very revealing of possible policy actions for congestion reduction. Empirical evidence confirms that trip-scheduling adaptation occurs during the morning commute, and that its derived effects are sometimes very important; e.g., "peak shifting" as a result of capacity expansions (Small, 1992).

Because of its appeal and simplicity, Vickrey's model has been extensively analyzed under different demand assumptions (Hendrickson et al., 1983; Kocur et al., 1984; Smith, 1984; Daganzo, 1985; Newell, 1987, Arnott et al., 1993a) and has also been adopted to analyze various toll policies (Arnott et al., 1990; Laih, 1994; Daganzo and Garcia, 2000). The model, however, only applies to cases where congestion is concentrated at a single location, affecting all commuters equally. These conditions are violated when the access network is itself congested. For example, freeway queues caused by bottlenecks often spill over long distances imposing different penalties on its access points. (Experimental evidence shows that congestion affects most severely origins far upstream of a bottleneck; Cassidy and Mauch, 2001). Obviously, network effects should be investigated.

Unfortunately, existing extensions of Vickrey's model to multi-origin/multi-destination networks (Kuwahara and Newell, 1987; Bernstein et al., 1993; Arnott et al, 1993b; Ran et al., 1996; Akamatsu et al., 1996) invariably ignore the spatial extent of queues (by

assuming *point-queues* that do not take up space) and oversimplify merging interactions.[1] These omissions may lead in general to wrong predictions, hence the insights derived for policy-making must be regarded with care.

This paper fills some of these gaps by introducing a network model that integrates Vickrey's theory with a realistic traffic flow model (Newell,1993) and a reasonable merging mechanism (Daganzo, 1996). Our ultimate goal is the qualitative understanding of the relationship among congestion, departure time choice and the spatial distribution of population for the morning commute, recognizing the networks are congested and have different origins. In this paper, we consider the simplest network with all the relevant characteristics. It consists of two origins, one destination and two links merging into a third, as shown on Figure 1. (This is the same network used in Arnott et al, 1993b to analyze departure-time equilibrium under the point-queue assumption.) Extensions to multi-origin networks will be considered in a sequel.

The paper is structured as follows. Section 1 introduces relevant background and discusses the single bottleneck (Vickrey) model. Section 2 presents the equilibrium model; first, basic concepts (Section 2.1), and then results (Section 2.2). It is found that merging and spillover effects affect equilibrium solutions significantly and in a rather counterintuitive way. Section 3 compares the solutions with those obtained under *point-queue* assumptions. Finally, section 4 discusses policy implications and relates them to earlier work.

---

[1] Kuwahara and Newell (1987) analyzes equilibrium on a population of commuters continuously distributed over space which must choose one (and only one) among several point-queue bottlenecks to reach a common destination. Friesz et al. (1994) describes a heuristic to solve network problem with point queues. Tests are done for a single origin-destination network. Akamatsu et al. (1999) considers general many-to-one/one-to-many networks under point-queues and constant *saturation* outflows.

1. The single bottleneck model

It is commonly assumed that traffic conditions during the morning commute are similar day-after-day. Commuters, aware of these, choose their departure time to minimize their individual trip cost, which consist of a trip-time component and a *schedule* penalty. The latter is associated with the arrival time at the destination relative to a *preferred* time. In the case where the only traffic restriction is a single bottleneck of capacity μ with no delay elsewhere, it is customary to express commuter decisions as a function of the *preferred passage time* through the bottleneck or *deadline*. If w is the deadline for the commuter that passes the bottleneck at time t and we express costs in units of trip time, then the trip cost for that commuter is

$$c = \tau + p(t - w) \tag{1}$$

where τ is the trip time and $p(.)$ is a schedule penalty function such that $p(.) \geq 0$ and $p(0)=0$. It will be assumed here that $p(.)$ is piecewise linear and V-shaped, where e and L are the positive conversion rates for *earliness* and *lateness* into trip time; i.e.,

$$p(s) = \begin{cases} -es & \text{if } s < 0 \\ Ls & \text{if } s \geq 0 \end{cases} \tag{2}$$

Normally, earliness is preferable to both queuing and lateness, i.e., $e<L$, $e<1$ (Small, 1982).

Vickrey's objective was to determine an equilibrium schedule of departures from a single origin such that no commuter/vehicle would have an incentive to change its departure time given the queues that resulted. The model also applies to multiple origins if all access routes to the bottleneck are uncongested and pass through a common point, O ; i.e., point O can be modeled as the single origin.

The solution can be represented by means of continuous cumulative plots, assuming that the number of commuters is so large that vehicles can be treated as a continuous variable; see Figure 2. $W(t)$ expresses the cumulative number of commuters wishing to pass the bottleneck by time t, and it will be called the *deadline curve*. It will be assumed that $W(t)$ is S-shaped, with slope greater than the capacity μ during some interval so that a queue must necessarily develop. $W(t)$ is a step function if all the commuters have the same deadline, as shown in Figure 2a. The objective is to finding an equilibrium curve of *cumulative arrivals* at the common point O, $A_O(t)$ – or equivalently the curve of cumulative *virtual arrivals* at the bottleneck, $A(t)=A_O(t-t_{OD})$ where t_{OD} is the fixed uncongested trip time from O to the bottleneck location, D.[2] According to standard queuing analysis, the curve of cumulative *departures* from the bottleneck, $D(t)$, is the highest curve with slope less than or equal to μ such that $D(t) \leq A(t)$.[3] Under a FIFO (first-in-first-out) queue, the delay τ for any given vehicle number is the horizontal distance between curves A and D for the given vehicle number. Likewise, the scheduled delay s is given by the horizontal distance between D and W if vehicles depart from the bottleneck in the order of their deadlines. It is known that if the penalty function $p(.)$ is convex and common to all commuters, the solution exists (Smith, 1985) and is unique (Daganzo, 1986). Furthermore, in the equilibrium solution, vehicles depart from the bottleneck in the order of their deadlines. An example of such equilibrium is represented in Figure 2 both for the case where commuters have a common deadline (Figure 2a) and

---

[2] A vehicle *virtual arrival* time to D is the time at which the vehicle would have passed D if it had travel unhindered from O to D.

[3] In queuing lingo, the terms *arrivals* and *departures* refer to the bottleneck. Therefore, they have the reverse meaning assigned to them in the economics literature where *arrivals to the bottleneck* correspond to the *departures from the origin* and vice-versa.

when they do not (Figure 2b). Both solutions exhibit a unique queuing episode with two clearly differentiated phases. In the first phase, commuters depart from the bottleneck earlier than desired and queuing delay increases with vehicle number at a rate that precisely compensates for the reduction in earliness. Therefore, the slope of $A(t)$ is given. In the second phase, commuters depart from the bottleneck later than desired and queuing time declines with vehicle number to compensate for increasing lateness penalties. As a result, the slope of $A(t)$ is also given. Note that the vehicle arriving on time experiences the highest delay (as given by the length of segment **AO**, |**AO**|, in Figure 2). In the single deadline case of Figure 2a, |**AO**| is the common cost suffered by all commuters.

If t_s and t_f are the times when the queue starts and vanishes, equilibrium requires $|\mathbf{AO}| = Lt_f = -et_s$. Furthermore, if we use N to denote the number of commuters who queue, then $N = \mu(t_f - t_s)$ since all these commuters depart when the bottleneck is at capacity. In the single deadline case, N is known (all the commuters queue), therefore these three equations define the three remaining unknowns: $|\mathbf{AO}|$, t_s and t_f. Since the slopes of $A(t)$ above and below **AO** are given, it follows that there is only one possible geometry for the equilibrium curves. Figure 2a shows that the number of commuters departing early at equilibrium is $N L/(e + L)$ and the number departing late is $N e/(e + L)$. One can also see that the common cost is $|\mathbf{AO}| = N eL/(e + L)$. Finally note that the equilibrium delay for a commuter departing at time t, $\tau(t)$, is:

$$\tau(t) = \begin{cases} |\mathbf{AO}| + et = e(t - t_s) & \text{if } t < 0 \\ |\mathbf{AO}| - Lt = L(t_f - t) & \text{if } t \geq 0 \end{cases} \tag{3}$$

which precisely balances the schedule penalty as required.

Consideration shows that a similar geometric pattern is an equilibrium for S-shaped deadline curves; see Figure 2b. The main difference is that in this case not all the commuters queue and therefore, one also needs to find N.

Most of the existing literature deals with fixed-capacity bottlenecks, but the analysis can be extended to variable capacities, $\mu(t)$, if $\mu(t)$ induces a single queuing episode.[4] Then, t_s, t_f and $A(t)$ can be determined as before since (3) continues to hold. This means that in any equilibrium, such as that shown in Figure 3a, the horizontal separation between A and D at $\# = D(t)$ continues to be given by (3). Therefore, if the equilibrium diagram is rescaled vertically by means of the transformation $\overline{\#} = D^{-1}(\#)$ which makes the departure rate equal to 1 at all times, i.e., $D^{-1}(D(t)) = t$, then we recover Figure 2a. This is shown in Figure 3b. The re-scaled arrival curve, $T(t) = D^{-1}(A(t))$, now returns the departure time t_d (on the vertical axis) as a function of the arrival time t_a (on the horizontal axis). We shall refer to $T(t)$ as the *arrival-departure schedule* curve (or *A/D* curve) to differentiate it from the equilibrium arrival curve, $A(t)$. The invariance of the rescaled diagram with respect to $\mu(t)$ will become useful later.

It should be remembered that the single bottleneck model does not apply if delays suffered by vehicles entering the network at different locations are different, as is normally the case for freeway networks. Unfortunately, no existing model addresses the three key effects required to model a simple freeway: multiple origins, merging interactions and queue spillovers. The next section describes a first step in this direction.

---

[4] This is always true if commuters have a common deadline.

2. Departure-time equilibrium with different origins and realistic traffic behavior

We shall consider here the simplest network exhibiting all three effects; see Figure 1. On this network, $N^{(A)}$ and $N^{(B)}$ commuters travel everyday from origins A and B to a common destination D. The routes from these origins merge at an intermediate location, M, and share a final link MD of length ℓ. A bottleneck of time-dependent capacity $\mu_D(t)$ may exist just upstream of D. Thus, queues may form on the common link and spill over the merge.[5]

Traffic is assumed to behave as in the *kinematic wave* (KW) theory. It will be modeled as in Newell (1993), where traffic obeys a triangular fundamental relationship linking flow, q, with density, k, as in Figure 4a. The relationship is defined by three parameters: a fixed free-flow speed (v_f), a maximum flow or capacity (μ) and a jam density (k_j).[6] We shall assume that the network is homogeneous (i.e. its three links have the same characteristics). The traffic model needs to be completed by defining how vehicles interact at the merge. We will use the rules in Daganzo (1996) as explained below. Finally, we shall assume that all commuters have the same deadline and penalty function (i.e., commuters are only distinguishable by their origin). Generalizations for networks with non-homogeneous links and different deadlines are discussed in Lago (2002).

2.1. Formulation

We express our equilibrium solution in terms of cumulative flows. As in Newell (1993), we shall use a *delay-based* formulation that ignores free-flow trip times. In this

---

[5] The flow restriction could be due to a variable inflow from another ramp (not depicted in the figure) very close to D.

[6] Jointly they define a wave speed (w) which represents the unique speed at which flow disturbances propagate upstream within a moving queue; see Figure 4.

formulation, a solution is defined in terms of the cumulative curves of *virtual arrivals* at point D $\{A^{(r)}(t), \ r = A, B\}$ – instead of the actual departure curves from origins A and B $\{A_r^{(r)}(t), \ r = A, B\}$ – and the cumulative curves of *departures* from D $\{D_D^{(r)}(t), \ r = A, B\}$.[7] Delays are given by $\tau^{(r)}(t) = t - A^{(r)^{-1}}(D_D^{(r)}(t)), \ r = A, B$. Newell (1993) shows that the *delay-based* traffic problem can be solved as a standard problem with the modified fundamental diagram of Figure 4b, which has the same μ and k_j but $v_f = \infty$. The actual departure curves from the origins can be obtained by shifting the virtual curves back in time by the origin-specific free-flow trip times (i.e., $A_r^{(r)}(t) = A^{(r)}(t + \ell_{rD}/v_f)$ where ℓ_{rD} is the distance from origin r to point D). Thus, from now on, and without loss of generality we take $v_f = \infty$ and use the diagrams of Figure 4b.

To solve the traffic-equilibrium problem, one could consider the arrival curves $\{A^{(r)}(t), \ r = A, B\}$ as the only unknowns since they define uniquely the cumulative departures from D $\{D_D^{(r)}(t), \ r = A, B\}$ and the delays for each origin $\{\tau^{(r)}(t), \ r = A, B\}$ through the traffic model. For our purposes, however, it will turn out to be more convenient to use $\{D_D^{(r)}(t), \tau^{(r)}(t), \ r = A, B\}$ as the unknowns and $\{A^{(r)}(t), \ r = A, B\}$ as the derived curves. Equilibrium conditions are more naturally expressed in terms of these unknowns as shown in section 2.1.1. On the other hand, since not every set of functions

---

[7] From now on, superscripts identify the origin to which the variable or function refers, while subscripts refer to the physical location over which the variable or function is defined; e.g., $q_D^{(r)}(t)$ is the flow at D of commuters from origin r. Furthermore, we will refer to $A^{(r)}(t)$ as *arrival curve* and $D_D^{(r)}(t)$ as *departure* curves following standard queuing notation.

$\{D_D^{(r)}(t), \tau^{(r)}(t), \ r = A, B\}$ is consistent with the traffic model, additional conditions must be specified. Fortunately, this is easy to do as explained in section 2.1.2.

Finally, for notational convenience, the origin-specific departure curves $\{D_D^{(r)}(t),$ $r = A, B\}$ will be expressed as a function of the *aggregated* departure curve, $D_D(t) = D_D^{(A)}(t) + D_D^{(B)}(t)$, and the proportion of departures by origin at time t, $\{\beta_D^{(r)}(t),$ $r = A, B\}$ so that $q_D^{(r)}(t) = \beta_D^{(r)}(t)q_D(t)$. We shall make use too of $D_M(t)$, the aggregated departure curve just downstream of merge M.

2.1.1. *Equilibrium conditions*

Equilibrium conditions have to be verified for each origin separately. The tuple $\{D_D(t), \beta_D^{(r)}(t), \tau^{(r)}(t)\}$, $r = A, B$ must be such that the trip cost for each origin r is equal for all the departure times from D with positive origin r outflows, $\beta^{(r)}(t) > 0$, and higher or equal elsewhere. Recall from (2) that the equilibrium delay $\tau(t)$ is affected by $D(t)$ through t_s and t_f only. Obviously the same simple principle applies now but with origin-specific $t_s^{(r)}$ and $t_f^{(r)}$. Thus if the departure processes $\{D_D(t), \beta_D^{(r)}(t), \ r = A, B\}$ are given and $\{t_s^{(r)}, t_f^{(r)}\}$ known, the equilibrium delays $\{\tau^{(r)}(t), \ r = A, B\}$ are given too. They in turn define the arrival processes $\{A^{(r)}(t), r = A, B\}$. If the traffic model generates the original departure processes from the equilibrium curves $\{A^{(r)}(t), r = A, B\}$, then the equilibrium is feasible. Thus, our problem is reduced to finding two departure processes $\{D_D(t), \beta_D^{(A)}(t)\}$ and $\{D_D(t), \beta_D^{(B)}(t)\}$ that define feasible equilibrium curves $\{A^{(A)}(t), A^{(B)}(t)\}$ and $\{\tau^{(A)}(t), \tau^{(B)}(t)\}$. The feasibility conditions imposed by the traffic model are presented below.

2.1.2.   Feasibility conditions

Our solutions should be consistent with both link and node dynamics.

(a) *Link dynamics: FIFO conditions.*

Since *KW* queues are FIFO, we must guarantee that commuters from different origins passing *M* at the same time incur the same delay on link *MD*.[8] Therefore, the common delays in that link needs to be predicted. We use Newell's method (Newell, 1993). First, a *capacity curve* at *M* , $D_{M+}(t)$, is defined from the departure curve at *D* by the shift,

$$D_{M+}(t) = D_D(t - \frac{k_j}{\mu}\ell) + k_j\ell \qquad (4)$$

The capacity curve sets an upper bound to the cumulative number of commuters that can pass *M* by time *t*; see Figure 4c. The actual cumulative curve of vehicles passing through *M, $D_M(t)$,* is the lower envelope of $D_{M+}(t)$ and the cumulative number of commuters who would have passed *M* in the absence of a queue.

It is convenient to express Newell's method in terms of delays and to index delays by time of departure. To this end, note that $D_{M+}(t) - D_D(t)$ is an upper bound for the length of the queues at *MD* at time *t*. Hence, the horizontal separation between D_{M+} and D_D for a given departure time, $v_{MD}(t)$, is also an upper bound for the delay on link *MD* for a given the history of departures. For this reason, we will call v_{MD} the maximum delays or *M-delays*. The actual delay in the link *MD* for a given departure time *t*, $\tau_{MD}(t)$, is the lesser of $v_{MD}(t)$ and the delay that would arise from an arrival curve at *M* obtained by ignoring the spillover effects but including the merge effects. Since we have two origins with

---

[8] FIFO is not an issue for links *AM* and *BM* because these links handle traffic from a unique origin.

known equilibrium delays $\tau^{(A)}(t)$ and $\tau^{(B)}(t)$, we can compare these delays with $v_{MD}(t)$ in an attempt to infer the actual delays $\tau_{MD}(t)$.

Since $0 \leq \tau_{MD} \leq v_{MD}$ and $0 \leq \tau_{MD} \leq \{\tau^{(A)}, \tau^{(B)}\}$, four cases can arise:

(i) $\quad 0 = v_{MD} = \tau_{MD} \leq \{\tau^{(A)}, \tau^{(B)}\}$ (i.e., delays only upstream of M)

(ii) $\quad 0 < v_{MD} = \tau_{MD} \leq \{\tau^{(A)}, \tau^{(B)}\}$ (i.e., delays upstream and downstream of M with

maximum delays on MD)

(iii) $\quad \tau_{MD} = \tau^{(A)} = \tau^{(B)} < v_{MD}$ (i.e., delays only downstream of M)

(iv) $\quad \tau_{MD} \leq \{\tau^{(A)}, \tau^{(B)}\} < v_{MD}$ (i.e., delays upstream and downstream of M but with less

than the maximum delay on MD).

We restrict our search to solutions where only (i), (ii) or (iii) occur, since this is the common case in most equilibrium solutions.[9] With this provision, $\tau_{MD}(t)$ can be expressed as a function of the data; i.e.,

$$\tau_{MD}(t) = \min\{v_{MD}(t), \max\{\tau^{(A)}(t), \tau^{(B)}(t)\}\}. \tag{S1}$$

Furthermore, the FIFO feasibility conditions can be expressed as:

$$\tau_{MD}(t) < v_{MD}(t) \Rightarrow \tau^{(A)}(t) = \tau^{(B)}(t) = \tau_{MD}(t) \tag{S2a}$$

$$\tau_{MD}(t) = v_{MD}(t) \Rightarrow \tau^{(A)}(t), \tau^{(B)}(t) \geq \tau_{MD}(t). \tag{S2b}$$

---

[9] Case (iv) does not arise as long as the capacity of the bottleneck $\mu_D(t)$ does not increase sharply before the deadline. When case (iv) arises, an equivalent method can be used based on a different (but more cumbersome) expression of $\tau_{MD}(t)$; see Lago (2002).

(b) *Node dynamics: merge rules*

Daganzo (1996) proposes that the discharge flows from the two upstream approaches of merge M, $\{q_M^{(A)}, q_M^{(B)}\}$, should be a function of the capacity of the upstream approaches (i.e., μ), the available capacity downstream (μ_M) and a priority ratio (α). For the case of interest here where $\mu_M \leq \mu$, the upstream approach capacity does not play a role and the rules reduce to the following two:

(i) During periods when there are no queues upstream of M, arrival flows equal discharge flows and $q_M = q_M^{(A)} + q_M^{(B)} \leq \mu_M$,

(ii) When there is a queue on approach r then $q_M = \mu_M$ and the departure ratio $\beta_M^{(r)} \equiv q_M^{(r)} / q_M \geq \alpha^{(r)}$, where $\alpha^{(A)} + \alpha^{(B)} = 1$.

It follows from (ii) that when there is a queue in both approaches, then $(q_M^{(A)} / q_M^{(B)}) = (\alpha^{(A)} / \alpha^{(B)})$.[10] These rules are illustrated in Figure 5. Recall that the downstream capacity may be time-dependent since $\mu_M(t) \equiv D'_{M+}(t)$.

Again, we need to express the merging rules in terms of our candidate departure and delay curves. Conditions (i) and (ii) imply that

$$D'_M(t) = q_M(t) \leq \mu_M(t) = D'_{M+}(t).$$ (M1)

Recall that $\tau^{(r)}(t) > v_{MD}(t)$ implies that vehicles departing at t from approach r experienced delay at the merge. Hence, in view of condition (ii), they entered the merge in a proportion $\beta_M^{(r)} \geq \alpha^{(r)}$. Since traffic on MD satisfies FIFO conditions, this proportion must be preserved at D (i.e., $\beta_M^{(r)}(t - \tau_{MD}(t)) = \beta_D^{(r)}(t)$) and we can write:

$$\tau^{(r)}(t) > v_{MD}(t) \Rightarrow \beta_D^{(r)}(t) \geq \alpha^{(r)} \qquad\qquad \text{(M2a)}$$

$$\tau^{(A)}(t) > v_{MD}(t), \tau^{(B)}(t) > v_{MD}(t) \Rightarrow (\beta_D^{(A)}(t), \beta_D^{(B)}(t)) = (\alpha^{(A)}, \alpha^{(B)}) . \qquad \text{(M2b)}$$

(M2a) defines the conditions when the merge is at capacity with one congested approach and (M2b) with two. When the merge is not saturated, any flow proportions are allowed.

2.2. Analysis

We consider the single deadline problem and analyze it in two phases: (i) cases with no bottleneck restriction at D ($\mu_D(t) = \mu$) where queues cannot form on link MD and merging effects dominate, and (ii) cases with time-dependent flow restrictions at D ($\mu_D(t) \leq \mu$) where queue spillovers can affect performance.

2.2.1. *No downstream restrictions* (*Merging effect*)

When no restrictions exist downstream of merge M, no delays can arise beyond it. Therefore, we can ignore link MD and treat M as if it was the destination (using $D_M(t) \equiv D_D(t)$, $\beta_M^{(r)}(t) \equiv \beta_D^{(r)}(t)$). In essence, the system is modeled as a pair of single-origin bottlenecks with departure rates coupled by the merging rule.

We shall look for equilibrium solutions in which the aggregated departure curve, $D_M(t)$, is given by the single bottleneck solution with capacity μ and total population $N^{(A)} + N^{(B)}$; i.e., where the merge is saturated only during the preferred interval $\Pi = [t_s, t_f]$ of Figure 2a. This is reasonable since the bottleneck mechanism allows undelayed travel from both origins when the merge is undersaturated (therefore, commuters would have an incentive to saturate the merge at the preferred times).

---

[10] A more detailed description of the dynamics of the merge section can be found in Daganzo, 1996.

Since D_M is given, the capacity shares $\{\beta^{(A)}(t), \beta^{(B)}(t)\}$ need to be found. If we assume

that only one queuing episode occurs on each approach, the solution is as shown in

Figure 6a. The figure displays the arrival and departure pattern on the two approaches

separately. Commuters from one origin (B in the figure) flow through the bottleneck

during an interval when the other approach is queued and therefore use a fixed share of

the capacity $\beta_M^{(B)}(t) = \alpha^{(B)}$. The solution for B-users is a single bottleneck equilibrium

with population $N^{(B)}$ and capacity $\alpha^{(B)}\mu$, see the bottom part of

Figure 6a, curves $D_M^{(B)}$ and $A^{(B)}$. Commuters from A flow at full capacity μ when the

approach B is not active and at a reduced capacity $\alpha^{(A)}\mu$ otherwise. The solution for A-

users is also a single bottleneck equilibrium, albeit with time-dependent capacity; see the

top part of

Figure 6a, curves $A^{(A)}$ and $D_M^{(A)}$.

The two diagrams of

Figure 6a can be re-scaled and superimposed following the procedure explained in

section 2 to show the *A/D curves* for both origins and the common departure curve on a

single diagram, see

Figure 6b.[11] The quantities in parenthesis following each colon are $\{\beta_D^{(A)}(t), \beta_D^{(B)}(t)\}$.

The curves on the figure must be similar, as shown, since commuters share a common

deadline and penalty function. Commuters from A experience the same commuting cost

---

[11] Remember that the horizontal distances between the *A/D curves* and D_M define the individual vehicle
delays on each approach for any departure time t.

as if everybody had the same origin, since **AO** is equal in length to the corresponding

segment of a single bottleneck solution with population $N^{(A)} + N^{(B)}$. Commuters from B,

however, experience a reduced cost, |**BO**|. From the figure, it is clear that B-commuters

experience less cost if $N^{(B)}/\alpha^{(B)} < N^{(A)}/\alpha^{(A)}$. The reverse is true if

$N^{(B)}/\alpha^{(B)} > N^{(A)}/\alpha^{(A)}$. The worst case arises if $N^{(B)}/\alpha^{(B)} = N^{(A)}/\alpha^{(A)}$ when all

commuters experience the highest cost. The best case arises if $\alpha^{(B)} = 0$ or 1 (complete

priority) when one of the origins experiences the least possible cost.

These results have an economic interpretation. Since the capacity of M is a scarce

resource, commuters impose onto each other an external cost (delay) as they jockey for

their preferred departure times. In the single bottleneck scenario, everybody is affected

equally by the actions of the others and the result of this game is a symmetric

equilibrium. A saturated merge, however, allocates its capacity in fixed shares – $\alpha^{(A)}$ and

$\alpha^{(B)}$ – to the two approaches, which insulates A-drivers from actions of B-drivers and

vice versa. This allows one part of the population to reduce its cost by traveling at the

most desired times. The effect can be exploited by manipulating ($\alpha^{(A)}, \alpha^{(B)}$). This shows

that *Pareto-improving* ramp metering schemes exist for single-destination freeways with

elastic demand.

2.2.2. *Downstream restrictions* (*Spillover effect*)

Let us now assume that $\mu_D(t) \leq \mu$, with the possibility of spillovers from link MD. We

shall find a feasible equilibrium in a sequential manner. Figure 7 illustrates the steps.

Step 1: Aggregate departure curve and maximum delay. We look for solutions in which

the aggregate departure curve, $D_D(t)$, is given by the single bottleneck solution with time-

variable capacity $\mu_D(t)$ and total population $N^{(A)} + N^{(B)}$ (see Figure 7a). This defines the duration of the queuing episode, $\Pi = [t_s, t_f]$. We shall assume that A-vehicles discharge uninterruptedly in the interval $\Pi^{(A)} = \Pi$ (i.e., $t_s^{(A)} = t_s$, $t_f^{(A)} = t_f$), and that B-vehicles discharge in an interval $\Pi^{(B)} \subset \Pi$ (i.e., $t_s^{(B)} \geq t_s^{(A)}$ and $t_f^{(B)} \leq t_f^{(A)}$), possibly with interruptions. The roles of the origins can be reversed, of course, and we show below that this depends on the *population-to-priority* ratio.

Since $t_s^{(A)}$ and $t_f^{(A)}$ are given, we can define an equilibrium delay function for origin A, $\tau^{(A)}$, as explained in section 2.1.2. This can be expressed graphically in conjunction with curve D_D by means of an *A/D curve* $T^{(A)}$ such that the horizontal difference between $T^{(A)}$ and D_D for any departure time t is the A-delay: $\tau^{(A)}(t) = t - T^{(A)^{-1}}(D_D(t))$; see Figure 7a. Neither D_D nor $T^{(A)}$ are cumulative counts for A-vehicles since D_D gives the cumulative count for A and B-vehicles together. Note too that the equilibrium delays for vehicles from origin B can never be larger than those of A since $\Pi^{(B)} \subset \Pi$, i.e., $\max\{\tau^{(A)}(t), \tau^{(B)}(t)\} = \tau^{(A)}(t)$.

Step 2: Delays on the common link. Since the aggregate departure curve D_D is given, we can shift it according to (4) to obtain the capacity curve at M, D_{M+}, and the *M-delays*, $\nu_{MD}(t)$; see Figure 7b. Since $\tau^{(A)}(t)$ are the maximum delays, we can use (S1) to obtain the actual delays on link MD: $\tau_{MD}(t) = \min\{\nu_{MD}(t), \tau^{(A)}(t)\}$. In other words, the actual delays are the least of the horizontal difference between D_{M+} and D_D and the horizontal difference between $T^{(A)}$ and D_D. Therefore, the curve of cumulative flows through M, D_M, is the lower envelope of D_{M+} and $T^{(A)}$ and the horizontal difference between D_M and D_D is

the actual delay τ_{MD}; see Figure 7c. These delays are highlighted by the shaded area of the figure.

Step 3: Solution for secondary origin (B). We first look for a starting time $t_s^{(B)} > t_s = t_s^{(A)}$. For each candidate $t_s^{(B)}$, the equilibrium delays $\tau^{(B)}(t)$ are given. Thus, we can again define a curve $T^{(B)}$ such that the horizontal difference between $T^{(B)}$ and D_D gives directly the equilibrium delay for origin B; see Figure 7d. Since $\tau^{(B)}(t)$ cannot be less than $\tau_{MD}(t)$, no departures from B can take place when the curve $T^{(B)}$ dips below D_M (non-shaded band of Figure 7d). Positive departures rates for B-vehicles can only occur on the shaded areas in Figure 7d. Note that all B-users experience a delay $\tau^{(A)} > \tau^{(B)} \geq v_{MD}(t)$, so that they cross the merge when both approaches have queues. Hence, they will always flow through the merge using a share $\alpha^{(B)}$ of the capacity.This means that $\beta_D^{(B)}(t) = \alpha^{(B)}$ in the shaded intervals and $\beta_D^{(B)}(t) = 0$ outside. The total number of B-vehicles passing through D is, therefore, the product of α and the vertical projection of the shaded areas. By changing the starting time $t_s^{(B)}$ we can ensure that the correct number of B-vehicles, $N^{(B)}$, is discharged. This is the equilibrium starting time. Figure 7e shows the final result. This completes the procedure since the arrival and departure curves for origin B are just rescaled versions of the delay curve $T^{(B)}$ and the departure curve D_D: $D_D^{(B)}(t) = \int_{-\infty}^{t} \beta_D^{(B)}(\varepsilon) D_D{}'(\varepsilon) d\varepsilon$ and $A_D^{(B)}(t) = \int_{-\infty}^{t} \beta_D^{(B)}(\varepsilon) T^{(B)}{}'(\varepsilon) d\varepsilon$. If $t_s^{(B)}$ has to be at a (later) point where $D_D < D_M$, the procedure still applies but $\beta_D^{(B)}(t) = 0$ until $T^{(B)} \geq D_M$. Similarly, if $t_f^{(B)}$ is at a time when $D_D < D_M$, again $\beta_D^{(B)}(t) = 0$ where $T^{(B)} < D_M$.

Step 4: Solution for primary origin (*A*). Since $\beta^{(A)}(t) = 1 - \beta^{(B)}(t)$, the departure and

arrival curve for origin *A* can be built by scaling curves D_D and $T^{(A)}$:

$$D_D^{(A)}(t) = \int_{-\infty}^{t} \beta_D^{(A)}(\varepsilon) D_D'(\varepsilon) d\varepsilon \text{ and } A_D^{(A)}(t) = \int_{-\infty}^{t} \beta_D^{(A)}(\varepsilon) T^{(A)}{}'(\varepsilon) d\varepsilon.$$ Note $\beta_D^{(A)}(t) = \alpha^{(A)}$

or 1 in agreement with our assumption that $\beta_D^{(A)}(t) > 0$ during Π. This completes the

solution.

Figure 8 shows the spatial evolution of the queues in Figure 7e for time intervals where

different queuing patterns arise. Thin black arrows represent the observed flows at the

critical sections; thick white narrows, the movements of the head and tail of the queues.

Note that the queuing state (iv) of section 2.1.2 never arises as assumed.

Three different solution types can arise depending on the populations $N^{(A)}, N^{(B)}$ and the

priority ratios $\alpha^{(A)}, \alpha^{(B)}$.

Solution 1A. If the number of *B*-vehicles is so small that $t_s^{(B)} > t_s^{(A)}$, the solution is as

illustrated by Figure 7e. In this case, *B*-commuters suffer less cost then *A*-commuters

since |**OB'**|<|**OA'**|. This solution arises if $0 < N^{(B)} / \alpha^{(B)} \leq (N^{(A)} + N^{(B)}) - N_U$, where N_U

is the total number of vehicles that find the merge uncongested; see Figure 7c.

Solution 1B. A symmetric solution to 1A where *A*-vehicles experience less cost is

obtained by interchanging the superscripts *A* and *B* and replacing $\alpha^{(B)}$ by $\alpha^{(A)}$. This

solution arises if $0 < N^{(A)} / \alpha^{(A)} \leq (N^{(A)} + N^{(B)}) - N_U$.

Solution 2. It is also possible to find an equilibrium for the remaining situations with

intermediate values of $N^{(B)} / N^{(A)}$. In these cases, $t_s^{(B)} = t_s^{(A)}$ and both origins share the

same cost |**OA'**|= |**OB'**|. Consideration shows that the solution is now as in Figure 7f. As required by the traffic model, $(\beta_D^{(A)}(t), \beta_D^{(B)}(t)) = (\alpha^{(A)}, \alpha^{(B)})$ in the intervals when both approaches are queued (shaded area) and $(\beta_D^{(A)}(t), \beta_D^{(B)}(t))$ is arbitrary in the middle period when M is below capacity (cross-hatched area). An equilibrium is reached for any $(\beta_D^{(A)}(t), \beta_D^{(B)}(t))$ that generates total discharges matching the populations $N^{(A)}$ and $N^{(B)}$. This solution arises if $(N^{(A)} + N^{(B)}) \geq N^{(B)} / \alpha^{(B)} \geq (N^{(A)} + N^{(B)}) - N_U$.[12]

A comparison of Figure 7e and Figure 7f with the single origin solution (Figure 3) and the solution with merging effects only (

Figure 6) reveals some interesting insights. In all cases, the population from the origin with the largest *population-to-priority ratio*, A, always experiences the same cost equivalent to the cost they would suffer in a single origin scenario with total population $N^{(A)} + N^{(B)}$. But commuters from the other origin, B, can incur lower cost. Curiously, this cost reduction decreases with the number of vehicles that can be stored in link MD ($k_j\ell$). In the extreme case where $k_j\ell \to 0$, we recover the solution of section 2.2.1 (albeit with a time-dependent capacity) which is actually the least total cost scenario. This is may seem paradoxical at first sight since the provision of extra storage space in link MD makes things worse even though the link capacities remain unchanged (!). The explanation is that the extra space allows A-commuters to mix with B-commuters in the downstream queue. This dilutes part of the segregation advantages that the merge gave to origin B. Again the policy-making implications of these effects are discussed in the final section.

---

[12] Note that since the value of N_U changes continuously with $(N^{(A)} + N^{(B)})$ and $k_j\ell$, a solution always exists.

3. Departure time equilibrium in point queue models with fixed link capacities

In the last decade, a number of works (Bernstein et al., 1993; Arnott et al., 1993b; Ran et al., 1996; Akamatsu et al., 1996) have proposed point-queue models with fixed link capacities to solve network equilibrium subject to departure-time choice. These models improve the realism of previous static models by including transient queuing phenomena, but they are still quite restrictive since they neglect two important forms of link-to-link interaction: *tail-to-head* (spillovers) and *head-to-head* (merging competition). These omissions lead to predictions that substantially overstate total cost as shown below. It is therefore important to incorporate link interactions in network models. The goal is achievable because the complexity of the problem appears not to be increased by doing so.

Consider first the issue of over-prediction and start with the constant-capacity case, $\mu_D(t) = \mu$. Since the above-mentioned point queue models do not restrict merging flows they allow queues to develop at D, although these queues would never appear in reality. The delays produced by these queues must be common to both origins. The ideas of section 2.2.2 can now be used to see that Figure 9b is a point-queue equilibrium with queues at D when $D_M'(t) > D_D'(t)$. This can be compared with Figure 9a, which is the solution of section 2.2.1 with $(\alpha^{(A)}, \alpha^{(B)}) = (0.5, 0.5)$. Clearly, the point-queue models predict a significantly larger equilibrium cost for B-users, $|\mathbf{OB}^{PQ}|$, and the wrong location and length of the queues. In the variable capacity case, the delay overstatement is exacerbated by the omission of spillovers as well (Lago, 2002).

Apart from their realism, an advantage of models with spillover and merge effects is that can be solved more easily. The biggest complication with all models (see Figure 9) is

keeping track of the common delays on link *MD*. But, with merges and spillovers, these delays turn out to be a function of the downstream conditions and the maximum delays, which are known; see condition (S1). This knowledge can be used to decouple the equilibrium solution by origin, and to simplify the solution method. (The decomposition is so effective that it can even be applied to most multi-origin networks; see Lago, 2002, and the sequel to this paper). The same simplification does not work with fixed-capacity, point-queue models, however, because in this case the delays on link *MD* always depend on the origin-specific flows from upstream. Thus, point-queue models are more difficult to generalize.

4. Policy implications

It was shown in Arnott et al. (1993b) that total system cost could decrease if one decreases the capacity of a link in a fixed-capacity point-queue network without a route choice. The result is interesting because it suggests that ramp-metering schemes could yield benefits in situations where the conventional wisdom (with fixed departure times) would say that none are to be had. On the other hand, the finding is of limited use because the assumptions underlying it rarely arise in ramp-metering practice. Thus, it is fair to ask whether the effect would also arise under the less restrictive assumptions of this paper, and whether it would be any more prevalent. The findings at the end of section 2.2.1 suggest that the answer to both questions should be affirmative—since it was shown that if storage capacity is not an issue then a Pareto-improvement can always be achieved by giving some priority to one of the origins. Although this result was only demonstrated for a network where both approach capacities were equal or greater than the capacity at the destination, the result is completely general. This is true because the

analysis for the general case turns out to be identical both in form and outcome to that of section 2.2.1 if the metering rate is constrained never to starve the destination bottleneck for flow. The beneficial effects of metering also arise in the time-dependent case with finite storage; see Lago (2002). The reason for the generality is that the merge allows the origin flows to interact in a detrimental way, and this happens whether or not the queue-mixing effect identified in Arnott et al (1993b) also arises. Since in most cases priority should go to the narrower approach, the results suggest that contrary to common practice priority in multi-origin freeways should go to the ramps closest to the bottleneck.

It was also shown in section 2.2.2 that reducing the storage capacity of link *MD* (i.e., its length) can reduce delay. This result is just as interesting because it shows that bringing the origins closer to the destination not only decreases free-flow travel time, but it also decreases delay (!). If the effect continues to arise with multi-origin networks, as we expect, it should have significant policy ramifications because it indicates that the travel costs added by congestion decrease with population density, if one holds the total population constant. This may seem paradoxical because it says that the denser a city the lesser its crowding cost. If such an effect turns out to be real, it should become part of the policy debate on urban sprawl.

It should be obvious from this discussion that further modeling efforts are needed if one is to develop a deeper appreciation for the effects of policy actions on congestion reduction. A reasonable second step, explored in the sequel, would extend the analysis to multi-origin networks with variable deadlines.

Acknowledgements

References

T. Akamatsu and M. Kuwahara, "A Capacity Increasing Paradox for a Dynamic Traffic Assignment with Departure Time Choice", in *Transportation and Traffic Theory* (*Proceedings of the 14th International Symposium*), A. Ceder (ed), 301-324, Pergamon Press, Amsterdam, 1999.

R. Arnott, A. De Palma, and R. Lindsey, "Economics of a bottleneck", *J. Urban Economics* **27**, 111-130 (1990).

R. Arnott, A. de Palma and R. Lindsey, "A structural model of peak-period congestion: A traffic bottleneck with elastic demand", *American Economic Review* **83**, 161-179 (1993).

R. Arnott, A. De Palma and R. Lindsey, "Properties of dynamic traffic equilibrium involving bottlenecks, including a paradox and metering", *Transportation Science* **27** (1993b).

D. Bernstein, T.L. Friesz, R.L. Tobin and B.W. Wie, "A Variational Control Formulation of the Simultaneous Route and Departure-Time Choice Equilibrium Problem", in *Transportation and Traffic Theory* (*Proceedings of the 12th International Symposium*), C.F. Daganzo (ed), 107-124, Elsevier, New York, 1993.

M.J. Cassidy and M. Mauch, "An observed traffic pattern in long freeway queues", *Transportation Research* **35A(2)**, 143-56 (2001).

C. F. Daganzo, "The uniqueness of a time-dependent equilibrium distribution of arrivals at a single bottleneck", *Transportation science* **19(1)**, 29-37 (1985).

C. F. Daganzo, "The nature of freeway gridlock and how to prevent it", in Transportation and Traffic Theory (Proceedings of the 13th International Symposium), J.B. Lesort (ed), 629-646, Pergamon Press, Amsterdam, 1996.

C. F. Daganzo and R. C. Garcia, "A Pareto improving strategy for the time-dependent morning commute problem", *Transportation science* **34(3)**, 303-311 (2000).

C. Hendrickson and G. Kocur, "Scheduled delay and Departure Time in a Deterministic Models", *Transportation science* **15(3)**, 62-77 (1981).

M. Kuwahara and G. F. Newell, "Queue Evolution on Freeways Leading to a Single Core City During the Morning Peak", in *Transportation and Traffic Theory* (*Proceedings of the 10th International Symposium*), N.H. Gartner and N.H.M. Wilson (eds), 21-40, Elsevier, New York, 1987.

A. Lago, Spatial models of morning commute consistent with traffic flow, Ph.D. thesis, University of California at Berkeley, Berkeley, CA, *to be published* 2002.

C.-H. Laih, "Queueing at a bottleneck with single and multi-step tolls", Transportation Research **28A(3)**, 197-208 (1994).

G. F. Newell, "The morning commute for non-identical travelers", *Transportation Science* **21(2)**, 74-88 (1987).

G. F. Newell, "A simplified theory of kinematic waves in highway traffic, I General Theory, II Queuing a freeway bottlenecks, III Multi-destination flows", *Transportation Research* **27B**, 281-313 (1993).

B. Ran, R.W. Hall and D.E. Boyce, "A Link-Based Variational Inequality Model for Dynamic Departure Time/Route Choice", *Transportation Research* **30B**, 31-46 (1996).

M. J. Smith, "The existence of a time-dependent equilibrium distribution of arrivals at a single bottleneck", *Transportation science* **18(4)**, 385-394 (1984).

K. A. Small, "The Scheduling of Consumer Activities: Work Trips", *American Economic Review* **72**, 467-479 (1982).

K.A. Small, "Trip Scheduling in Urban Transportation Analysis (in Transportation Economics)", *American Economic Review* **82(2)**, 482-486 (1992).

W. S. Vickrey, "Congestion Theory and Transport Investment", *American Economic Review* **59**, 251-260 (1969).

Figures

Figure 1. A simple network: 2 origins – 1 destination with merging routes.
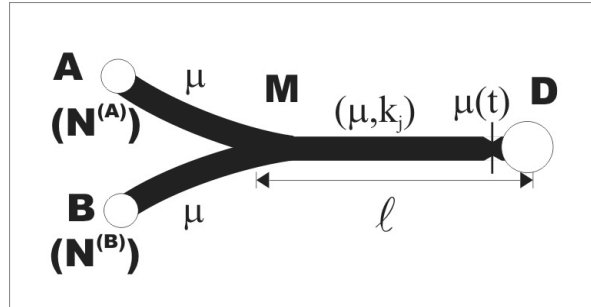
Figure 2. Equilibrium solution of the single bottleneck problem with fixed capacity.
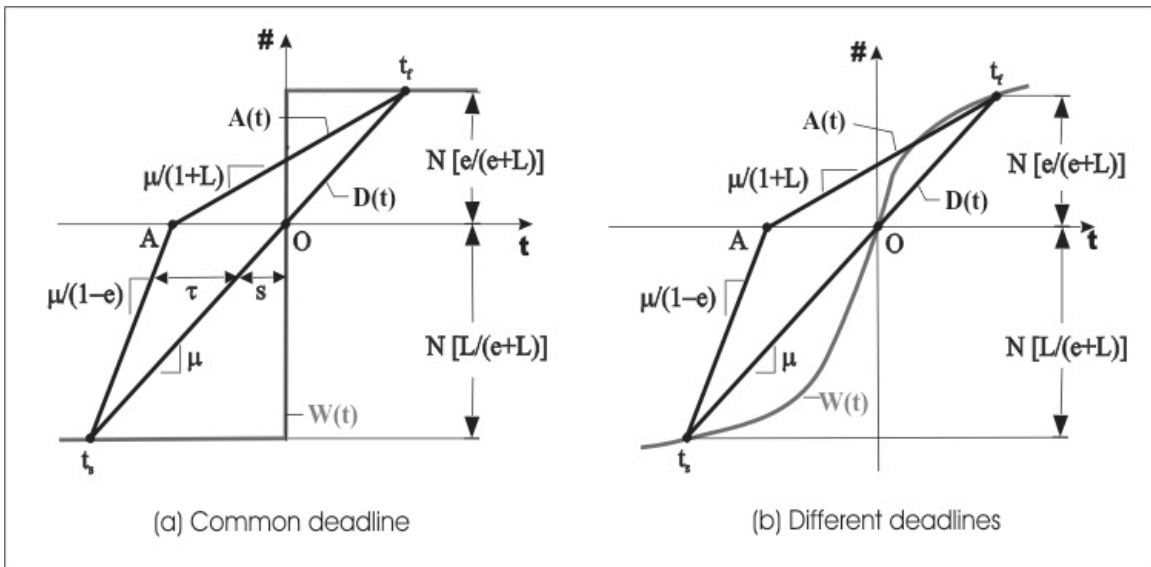
Figure 3. Equilibrium solution of the single bottleneck problem with variable capacity.
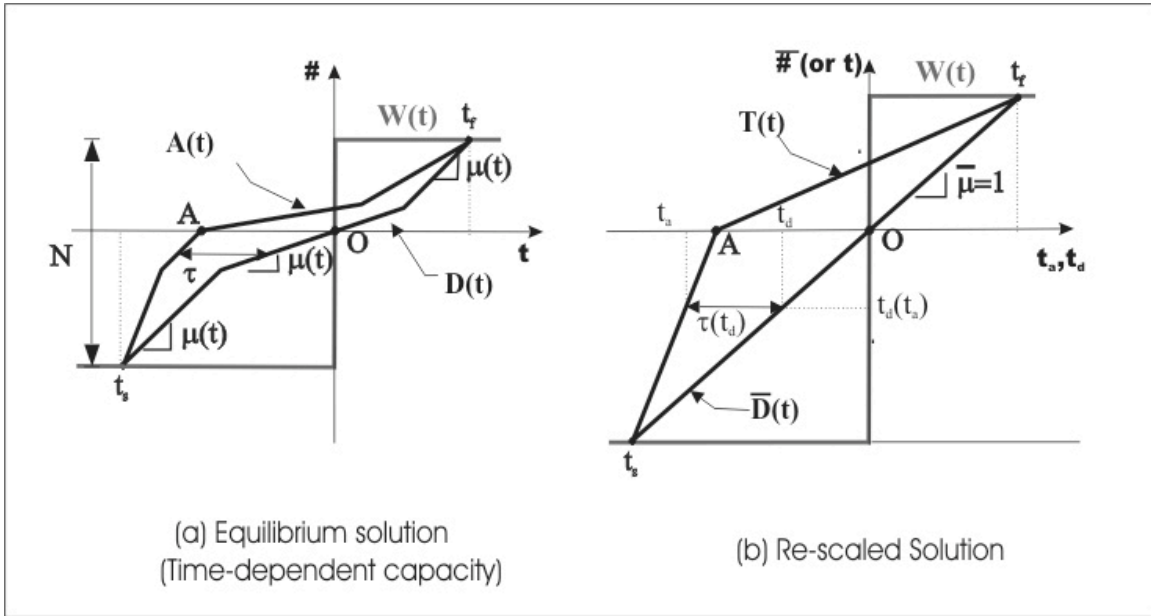
Figure 4. Traffic flow model.
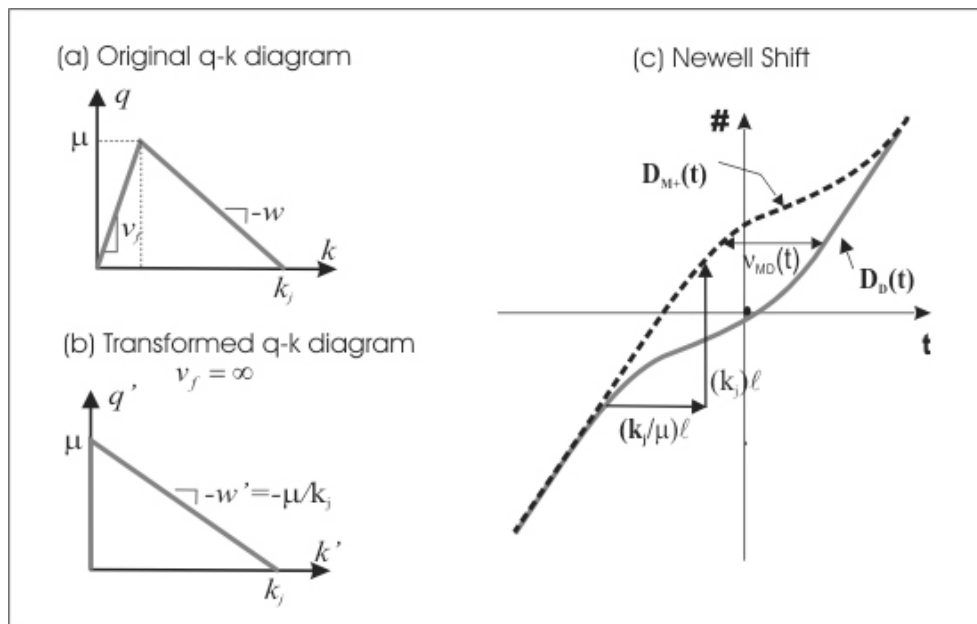
Figure 5. Merge diagram.
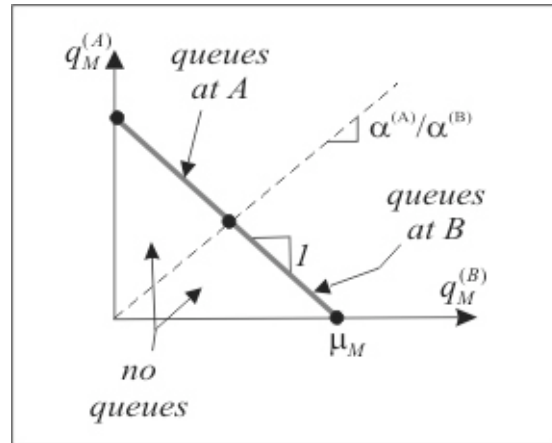
Figure 6. Equilibrium solution without queues in link *MD*.

Figure 7. Equilibrium solution with queues in link *MD*.

Figure 8. Physical queue evolution in equilibrium solution (type 1A). Sequence numbers

   refer to states shown in the cumulative plot.

Figure 9. Comparison of solutions under the KW model and the point-queue model, no

   downstream restrictions.

Figure 1. A simple network: 2 origins – 1 destination with merging routes.

Figure 2. Equilibrium solution of the single bottleneck problem with fixed capacity.

(a) Common deadline

(b) Different deadlines

Figure 3. Equilibrium solution of the single bottleneck problem with variable capacity.

(a) Equilibrium solution
(Time-dependent capacity)

(b) Re-scaled Solution

Figure 4. Traffic flow model.

(a) Original q-k diagram

(b) Transformed q-k diagram

(c) Newell Shift

Figure 5. Merge diagram.

Figure 6. Equilibrium solution without queues in link *MD*.

(a) Origin-specific solution   (b) Aggregated re-scaled solution

Figure 7. Equilibrium solution with queues in link *MD*.

Figure 8. Physical queue evolution in equilibrium solution (type 1A). Sequence numbers refer to states shown in the cumulative plot.
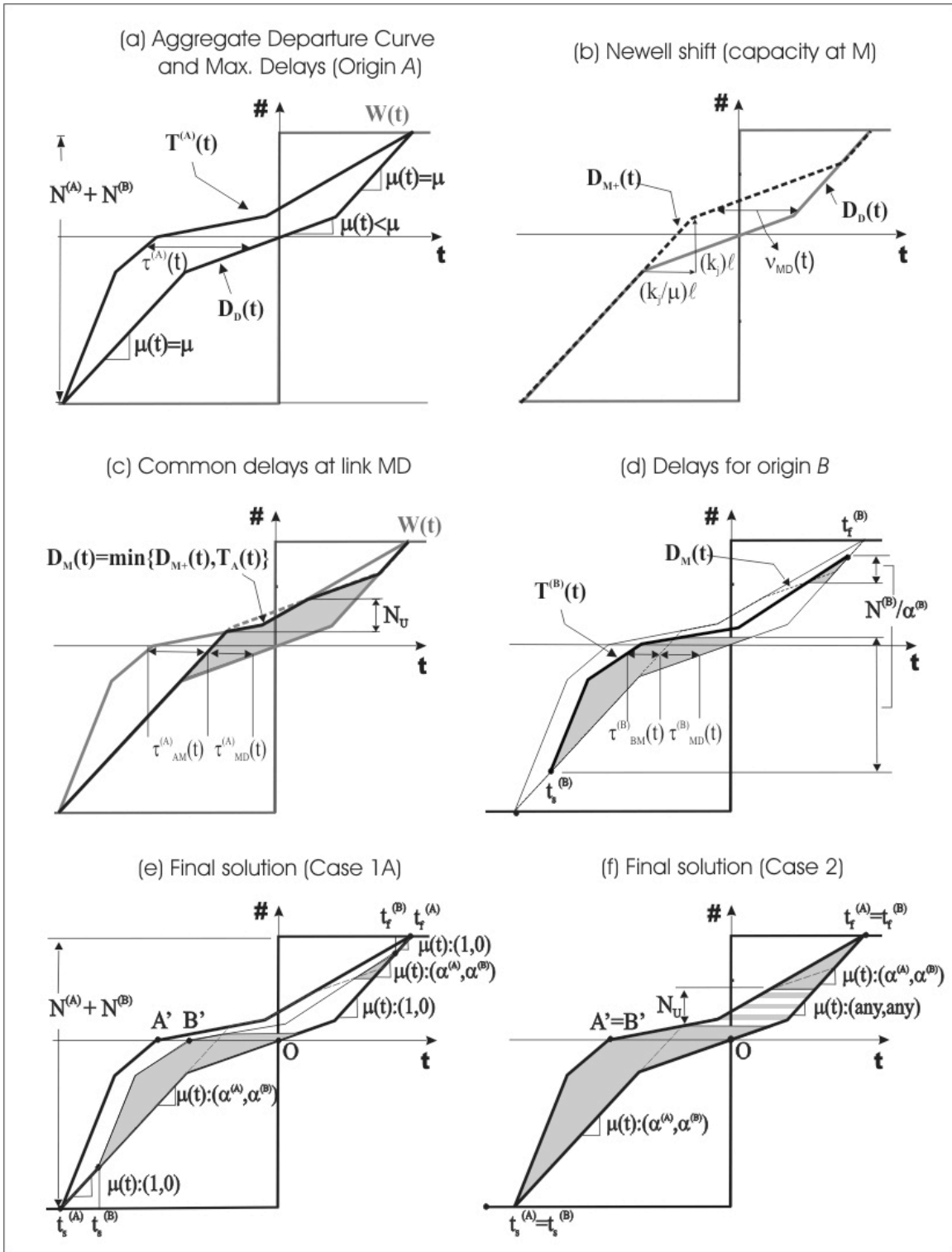
Figure 9. Comparison of solutions under the KW model and the point-queue model. No downstream restrictions.

(a) KW Equilibrium solution
(b) Point-queue Equilibrium solution