Dynamic user optimal assignment with physical queues for a many-to-many OD pattern

Masao Kuwahara a,*, Takashi Akamatsu b

a Institute of Industrial Science, University of Tokyo, 7-22-1, Roppongi, Minato-ku, Tokyo 106, Japan
b Toyohashi University of Technology, 1-1, Hibarigaoka Tenpaku-cho, Toyohashi-shi, Aichi 441, Japan

Received 10 October 1999; received in revised form 29 November 1999; accepted 6 December 1999

Abstract

This research extends the dynamic user optimal assignment under the point queue concept so as to deal with physical queues. Given time-dependent many-to-many OD volumes, this paper first shows the formulation of the assignment subject to the flow conservation and the first-in-first-out (FIFO) queue discipline. The optimal condition is then defined and the physical queue propagation based on the kinematic wave theory is discussed. Finally, a solution algorithm is proposed and typical differences between point and physical queue analyses are presented through an example calculation. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Dynamic assignment; Queueing delay; Dynamic user optimal; Physical queue

1. Introduction

Travel time delay is the major factor which controls users’ route choices and delay caused in a queue is normally much larger than delay outside a queue. Since queueing phenomena are the accumulation of vehicles over time which arrive at a bottleneck exceeding its capacity, the time-dependent analysis is required to incorporate queueing phenomena in traffic assignment. In most of previous studies in dynamic assignment, queues are treated as point queues or vertical queues which have no physical lengths, since point queues are robust to evaluate delay of any queueing system just from the queue length and service rate at the bottleneck of the system.

However, one of the shortcomings in point queues is that queue propagation phenomena cannot be described because of no physical length. For example, at a highway merging section, when a queue

*Corresponding author. Tel.: +81-3-3402-6231, ext. 25; fax: +81-3-3401-6286.
E-mail address: kuwahara@nishi.iis.u-tokyo.ac.jp (M. Kuwahara).

0191-2615/01/$ - see front matter © 2001 Elsevier Science Ltd. All rights reserved.
PII: S0191-2615(00)00005-9
fully occupies a freeway mainline, a vehicle from an on-ramp can sneak into the middle of the queue. In this case, since vehicles from the on-ramp do not have to follow the tail of the physical queue, their delay would be considerably different from one in the point queue analysis, and consequently their route choices may be also different. The physical queue analysis would give more realistic assignment result especially when physical queues back up merging and/or diverging sections.

Newell (1993a,b) analyzes physical queues in relation to the cumulative arrivals and departures given a time-dependent OD table between ramps along one freeway. He employs the kinematic wave theory by Lighthill and Whitham and shows how to draw cumulative curves of physical queues at any location along a road section when the wave speeds and the jam density associated with the section are given.

Regarding network assignment problems, Daganzo (1994,1995) proposes the cell transmission model which reproduces traffic condition on a network with physical queues provided that routes of every OD pair are given. This is a sort of traffic simulation following the kinematic wave theory and applicable to a general network. For a simple network, Daganzo (1998) points out interesting effects of physical queues on the dynamic assignment problem in contrast with point queues. In particular, he shows that, in some situation, an increased capacity for one of the links can lead to a reduced output and to oversaturation. Heydecker and Addison (1996) studied the dynamic user equilibrium (DUE) assignment, in which users choose their routes based on actual experienced travel times (or costs), under the physical queue concept. They incorporate physical queues into the assignment employing the above theory by Newell, but the network is quite limited to one with single destination and no interaction among route flows (non-overlapping routes).

As in above, the theory on the dynamic assignment has not been well established for a general network with physical queues. For a general network but under the point queue concept, Kuwahara and Akamatsu (1997) previously proposed the formulations and decomposition of the dynamic user optimal (DUO) assignment with a many-to-many OD pattern. The DUO assignment is sometimes called the reactive assignment because vehicles are assumed to choose their routes based on present instantaneous travel times, which is different from the DUE assignment principle. This research extends the DUO assignment so as to deal with physical queues in order to analyze the queue backing-up phenomena on a general network. For the extension, our previous point queue analysis is combined with the theory by Newell mentioned above.

2. Definition of dynamic user optimal (DUO) assignment

This section briefly defines the DUO assignment. First, our network and traffic demand are introduced, and then the DUO principle is explained together with required constraints such as the FIFO discipline and the flow conservation based on our previous study (see Kuwahara and Akamatsu 1993, and Akamatsu and Kuwahara, 1994)).

2.1. Network and traffic demand

A network consists of links and nodes. Sequential numbers from 1 to N are allocated to N nodes. The number of links is L and a link from node i to j is denoted as link (i,j). A time-dependent many-to-many OD demand is assumed to be given, which is denoted as
\[ Q_{id}(t) = \text{cumulative OD demand from origin } i \text{ to destination } d \text{ generated at the origin by time } t \text{ (given)}. \]  

(1)

2.2. Constraints to be satisfied

2.2.1. Flow conservation at nodes

The flow conservation at a node must always be satisfied, which is written with respect to node \( i \) by introducing cumulative arrivals and departures of vehicles to a particular destination \( d \):

\[ -\sum_h t_{hi}^d(t) + \sum_j A_{ij}^d(t) = Q_{id}(t) \quad \forall i, \forall d, \ i \neq d, \]  

(2)

where

\[ A_{ij}^t(t) = \text{the cumulative arrivals at link } (i, j) \text{ to destination } d \text{ by time } t, \]

\[ D_{ij}^t(t) = \text{the cumulative departures from link } (i, j) \text{ to destination } d \text{ by time } t. \]

The derivative of the conservation equation yields

\[ -\sum_h t_{hi}^d(t) + \sum_j \lambda_{ij}^d(t) = q_{id}(t) \quad \forall i, \forall d, \ i \neq d, \]  

(3)

where \( \lambda_{ij}^d(t) = \frac{dA_{ij}^d(t)}{dt}, \ \mu_{ij}^d(t) = \frac{dD_{ij}^d(t)}{dt}, \ q_{id}(t) = \frac{dQ_{id}(t)}{dt}. \)

The arrival rate at link \( (i, j) \) to destination \( d \) at time \( t \), \( \lambda_{ij}^d(t) \), is the unknown variable which must be determined so as to establish the DUO assignment principle. The cumulative arrival, \( A_{ij}^d(t) \), is the integral of \( \lambda_{ij}^d(t) \) over time until time \( t \), and our objective is thus to determine \( A_{ij}^d(t) \) at every link for any \( t \).

2.2.2. First-in-first-out discipline

Under the FIFO discipline, vehicles must leave \( (i, j) \) in the same order as the order of arrivals at the link. Thus, the \( A_{ij}(t) \) and \( D_{ij}(t) \) defined below must be related to each other through actually experienced link travel time \( T_{ij}(t) \) irrespective of destination \( d \):

\[ A_{ij}(t) = D_{ij}(t + T_{ij}(t)), \]  

(4)

where \( T_{ij}(t) \) is the experienced travel time on link \( (i, j) \) for a vehicle entering the link at time \( t \):

\[ A_{ij}(t) = \text{the cumulative arrivals at link } (i, j) \text{ by time } t = \sum_d A_{ij}^d(t), \]

\[ D_{ij}(t) = \text{the cumulative departures from link } (i, j) \text{ by time } t = \sum_d D_{ij}^d(t). \]

This condition must also be satisfied even by vehicles traveling toward the same destination node \( d \): \( A_{ij}^d(t) = D_{ij}^d(t + T_{ij}(t)). \) And by taking derivative with respect to time \( t \), the FIFO discipline is described in slightly different form as follows:

\[ \frac{\lambda_{ij}^d(t)}{\lambda_{ij}(t)} = \frac{\mu_{ij}^d(t + T_{ij}(t))}{\mu_{ij}(t + T_{ij}(t))}. \]  

(5)
3. Physical queues

3.1. Modification of the cumulative curves based on the kinematic theory

Each link has a certain value of its flow capacity and a queue forms when traffic demand exceeds its capacity. Suppose a network illustrated in Fig. 1, in which the downstream link has less capacity than the upstream link due to the lane reduction. Though a point queue vertically forms at the exit point of the upstream link (the capacity bottleneck), a physical queue forms horizontally. And once the link is fully occupied by the physical queue, the departure flow rate from the further upstream link is also restricted to the capacity of the bottleneck.

To determine how fast a physical queue propagates backward, the wave speed is analyzed by introducing the following new variable:

\[ F_{ij}(x, t) = \text{the cumulative number of vehicles passing at location } x \text{ on link } (i, j) \text{ by time } t, \]

where location \( x \) is measured from the upstream end of a link toward downstream (same as the flow direction).

By definition, flow \( f_{ij}(x, t) \) and density \( k_{ij}(x, t) \) of the link are, respectively, given by \( f_{ij}(x, t) = \partial F_{ij}(x, t)/\partial t \) and \( k_{ij}(x, t) = -\partial F_{ij}(x, t)/\partial x \). The cumulative arrival and departure defined in the previous section are also described: \( D_{ij}(t) = F_{ij}(\ell_{ij}, t) \) and \( A_{ij}(t) = F_{ij}(0, t) \).

Considering the kinematic wave theory that flow does not change on the trajectory of a wave (characteristic curve), Newell (1993a,b) analyzes a relationship between the cumulative curves and the wave speed. In particular, when the flow-density relationship is drawn by only two wave speeds \( w_{ij} \) and \( -w'_{ij} \) as shown in Fig. 2 at any time \( t \) and location \( x \) on link \( (i, j) \), he shows that \( dF_{ij}(x, t)/dx \) takes the constant value of \( -k_{ij}^{\max} \) or 0 independent of location \( x \):

\[
\frac{dF_{ij}(x, t)}{dx} = \begin{cases} 
0 & \text{for the forward wave,} \\
-k_{ij}^{\max} & \text{for the backward wave.}
\end{cases}
\]  

(6)

As shown in Fig. 3, suppose that the cumulative number of vehicles entering link \( (i, j) \), \( A_{ij}(t) \), propagates forward with the wave speed of \( w_{ij} \). The cumulative arrivals at the downstream end, \( A_{ij}(t - \ell_{ij}/w_{ij}) \), is obtained by horizontally shifting \( A_{ij}(t) \) along the time axis by \( \ell_{ij}/w_{ij} \). But \( A_{ij}(t) \) is not shifted vertically, because \( dF_{ij}(\ell_{ij}, t)/dx = dA_{ij}(t)/dx = 0 \) for the forward wave from (6). If the slope of \( A_{ij}(t - \ell_{ij}/w_{ij}) \) exceeds the maximum service rate of link \( (i, j) \) at time \( t_0 \), a queue forms and the backward wave starts moving. Since \( dF_{ij}(0, t)/dx = dD_{ij}(t)/dx = -k_{ij}^{\max} \) for the backward wave, we construct the cumulative curve on the backward wave reaching the entry point of the

![Fig. 1. Difference in queueing between physical and point queues.](image-url)
link by shifting $D_{ij}(t)$ horizontally by $\ell_{ij}/w_{ij}'$ and vertically by $k_{ij}^{\text{max}} \cdot \ell_{ij}$; that is, $D_{ij}(t - \ell_{ij}/w_{ij}') + k_{ij}^{\text{max}} \cdot \ell_{ij}$. The intersection with $A_{ij}(t)$ indicates that the shock reaches the entry point. According to the theory, the lower line of $A_{ij}(t)$ or $D_{ij}(t - \ell_{ij}/w_{ij}') + k_{ij}^{\text{max}} \cdot \ell_{ij}$ shows the cumulative number of physical vehicles arriving at the entry point of link $(i,j)$. 

Fig. 2. A flow-density relationship on link $(i,j)$.

Fig. 3. Three-dimensional illustration of wave propagation on link $(i,j)$. 

flow $f_{ij}(x,t)$

[veh/unit time]

Density $k_{ij}(x,t)$

[veh/unit length]
3.2. Instantaneous link travel time

The instantaneous link travel time on link \((i, j)\) at time \(t\), \(\tau_{ij}(t)\), is defined as

\[
\tau_{ij}(t) = \int_0^{x_{ij}^*(t)} \frac{dx}{v_{ij}(x, t)},
\]

where \(v_{ij}(x, t)\) is the velocity of link \((i, j)\) at location \(x\) and time \(t\).

Note that instantaneous link travel time \(\tau_{ij}(t)\) is different from actually experienced travel time \(T_{ij}(t)\) defined in (4). Given the flow-density relationship as well as \(A_{ij}(t)\) and \(D_{ij}(t)\) by time \(t\), velocity \(v_{ij}(x, t)\) can be in principle evaluated at any location \(x\) on the link at time \(t\) (Newell, 1993a). First of all, the location \(x_{ij}^*(t)\) where a shock reaches at time \(t\) is determined such that the forward and backward waves coincide each other. Since the backward wave travels \(-\ell_{ij} - xi_j^*(t)\) at speed of \(-w_{ij}'\) and the forward wave travels \(x_{ij}^*(t)\) at speed of \(w_{ij}\), the location \(x_{ij}^*(t)\) is determined so as to satisfy the following equation:

\[
A_{ij}\left(t - \frac{x_{ij}^*(t)}{w_{ij}}\right) = D_{ij}\left(t - \frac{\ell_{ij} - x_{ij}^*(t)}{w_{ij}'}\right) + k_{ij}^{\text{max}} \cdot (\ell_{ij} - x_{ij}^*(t)).
\]

Clearly, location \(x_{ij}^*(t)\) can be determined if \(A_{ij}(t)\) and \(D_{ij}(t)\) are known by present time \(t\), although an iterative method is required to determine \(x_{ij}^*(t)\) from (8). If we take derivative (8) with respect to \(x_{ij}^*(t)\), the following are obtained:

\[
k_{ij}^{\text{max}} - \left\{ \lambda_{ij}\left(t - \frac{x_{ij}^*(t)}{w_{ij}}\right) \right\} / w_{ij} + \mu_{ij}\left(t - \frac{\ell_{ij} - x_{ij}^*(t)}{w_{ij}'}\right) / w_{ij}' > 0.
\]

The above derivative is always positive, if \(\lambda_{ij}()\) and \(\mu_{ij}()\) are bounded by \(J^{\text{max}}\). Therefore, \(x_{ij}^*(t)\) can be uniquely determined.

For \(x \geq x_{ij}^*(t)\), the traffic condition is in the congested region and the flow rate at location \(x\) at time \(t\) must be equal to the slope of

\[
D_{ij}(t - (I_{ij} - x)/w_{ij}') : \mu_{ij}(t - (I_{ij} - x)/w_{ij}).
\]

From the flow-density relationship, velocity \(v_{ij}(x, t)\) can be known as a function of the flow rate in the congested region

\[
v_{ij}(x, t) = v_{ij}\left(\mu_{ij}\left(t - \frac{\ell_{ij} - x}{w_{ij}'}\right)\right).
\]

In particular, if we employ the piece-wise flow-density relationship as shown in Fig. 2, the functional form of (9) is

\[
v_{ij}(x, t) = v_{ij}\left(\mu_{ij}\left(t - \frac{\ell_{ij} - x}{w_{ij}'}\right)\right) = \frac{\mu_{ij}(t - (\ell_{ij} - x)/w_{ij}')}{k_{ij}^{\text{max}} - \mu_{ij}(t - (\ell_{ij} - x)/w_{ij}')/w_{ij}'}, \quad x \geq x_{ij}^*(t).
\]
On the other hand, for \( x < x^*_j(t) \), the traffic condition is in the free flow region and the velocity is equal to free flow speed \( w_{ij} \). Therefore, the instantaneous link travel time on link \((i, j)\) at time \( t \) is

\[
\tau_{ij}(t) = \int_0^{\ell_{ij}} \frac{dx}{v_{ij}(x, t)} = x^*_j(t) + \int_{x^*_j(t)}^{\ell_{ij}} \frac{dx}{w_{ij}} \frac{1}{\mu_{ij}(t - (\ell_{ij} - x)/w_{ij})}.
\]

In this way, instantaneous link travel time \( \tau_{ij}(t) \) can be evaluated from \( A_{ij}(t) \) and \( D_{ij}(t) \) by present time \( t \).

### 3.3. Evaluation of departure flow rates

Under the point queue concept, the departure flow rate from link \((i, j)\) can be simply determined; that is, if the link has a queue, departure flow rate \( \mu_{ij}(t) \) is restricted to its constant link capacity, but otherwise \( \mu_{ij}(t) \) is equal to the arrival rate at \( \ell_{ij}/w_{ij} \) before, \( \lambda_{ij}(t - \ell_{ij}/w_{ij}) \). On the other hand, for a physical queue, the link capacity may change over time since a queue physically backs up and may reduce the link capacity of the upstream link such as for \( t_1 \) to \( t_2 \) in Fig. 3. The following discussion shows typical specifications of departure flow rates for simple, diverging and merging sections under the physical queue concept.

For a simple highway section such as upstream link \((i, j)\) followed by downstream link \((j, k)\), let us first define back-up-flag \( \beta_{jk}(t) \) which indicates queue backing up conditions on link \((j, k)\) at time \( t \):

\[
\beta_{jk}(t) = \begin{cases} 
1 & \text{if } A_{jk}(t) > D_{jk}(t - \ell_{jk}/w'_{jk}) + k_{jk}^{\text{max}} \times \ell_{jk} \text{ (queue backing up)}, \\
0 & \text{otherwise}.
\end{cases}
\]

Using \( \beta_{jk}(t) \), the possible acceptable flow rate of downstream link \((j, k)\), \( Y_{jk}(t) \), is written as

\[
Y_{jk}(t) = \{1 - \beta_{jk}(t)\} \times f_{jk}^{\text{max}} + \beta_{jk}(t) \times \mu_{jk}(t - \ell_{jk}/w'_{jk}),
\]

where \( f_{jk}^{\text{max}} \) is the maximum flow rate observed in the flow-density relationship on link \((j, k)\) as shown in Fig. 2. If a queue does not back up from downstream link \((j, k)\); \( \beta_{jk}(t) = 0 \), the \( Y_{jk}(t) \) is equal to maximum flow rate \( f_{jk}^{\text{max}} \), but when the shock reaches entry node \( j \), the \( Y_{jk}(t) \) would be reduced to the departure flow rate of the downstream link, \( \mu_{jk}(t - \ell_{jk}/w'_{jk}) \).

On the other hand, traffic demand which would like to leave link \((i, j)\) at time \( t \), \( X_{ij}(t) \), is written as

\[
X_{ij}(t) = \begin{cases} 
f_{ij}^{\text{max}}, & A_{ij}(t - \ell_{ij}/w_{ij}) > D_{ij}(t) \text{ or } \lambda_{ij}(t - \ell_{ij}/w_{ij}) > f_{ij}^{\text{max}}, \\
\ell_{ij}(t - \ell_{ij}/w_{ij}), & \text{otherwise}.
\end{cases}
\]

This means that if a queue exists on link \((i, j)\) at time \( t \), at most \( f_{ij}^{\text{max}} \) can leave link \((i, j)\), but otherwise, demand of \( \lambda_{ij}(t - \ell_{ij}/w_{ij}) \) would like to leave. The \( X_{ij}(t) \) can be recognized as the demand rate from link \((i, j)\) without considering the downstream condition.

Departure flow rate \( \mu_{ij}(t) \) is then written as the minimum of upstream demand rate \( X_{ij}(t) \) and downstream acceptable flow rate \( Y_{jk}(t) \):
\[ \mu_{ij}(t) = \text{Min}[X_{ij}(t), Y_{jk}(t)]. \]  

(15)

When node \( i \) is an origin node, OD demand \( q_{il}(t) \) is given but the entire generated demand rate from the origin, \( q_{il}(t) \), may not be able to enter the network if a queue backs up to the origin node. To maintain the flow conservation at origin node \( i \), we consider point queues only at origin nodes and a part of \( q_{il}(t) \) which cannot enter the network due to the queue backing up is accumulated as a point queue at origin node \( i \).

For merging and diverging sections, we assume the following for the simplicity:

I. Merging and diverging do not occur at the same node.

II. No demand is generated as well as attracted at any merging or diverging node.

These assumptions may not be unrealistic and, without loss of generality, we could code a network so as to meet these assumptions.

For a diverging section at node \( j \), provided that there is only one link \((i,j)\) entering the diverging node \( j \) as in assumption I, the departure flow rate would be determined as follows:

\[ \mu_{ij}(t) = \text{Min} \left[ X_{ij}(t), \sum_k Y_{jk}(t), \text{Min}_k \{ Y_{jk}(t) / \phi_{jk}(t) \} \right], \]

(16)

where \( \phi_{jk}(t) = \text{the ratio of diverging flows at node } j \) (\( = \lambda_{jk}(t) / \sum_{k'} \lambda_{k'}(t) \)) determined through the DUO assignment.

The third term of the right-hand side means that individual arrival rate \( \lambda_{jk}(t) \) should not exceed \( Y_{jk}(t) \) for \( \forall k \). Since the arrival rate \( \lambda_{jk}(t) \) is written as \( \phi_{jk}(t) \cdot \mu_{ij}(t) \), the constraint is derived as follows:

\[ \lambda_{jk}(t) = \phi_{jk}(t) \cdot \mu_{ij}(t) \leq Y_{jk}(t), \]
\[ \therefore \mu_{ij}(t) \leq Y_{jk}(t) / \phi_{jk}(t) \quad \text{for } \forall k. \]

This condition must be satisfied for each of the downstream link \((j,k)\)'s. Therefore, \( \mu_{ij}(t) \) is constrained by \( \text{Min}_k \{ Y_{jk}(t) / \phi_{jk}(t) \} \).

The diverging ratio \( \phi_{jk}(t) \) would be determined as below under assumptions I and II above as well as the FIFO discipline (5):

\[ \phi_{jk}(t) = \frac{\lambda_{jk}(t)}{\sum_{k'} \lambda_{k'}(t)} = \frac{\sum_d \delta_{jk}(t) \mu_{ij}(t)}{\sum_{k'} \sum_d \delta_{jk'}(t) \mu_{ij}(t)} = \frac{\sum_d \delta_{jk}(t) (\mu_{ij}(t) / \lambda_{ij}(t)) \cdot \lambda_{ij}(t)}{\sum_{k'} \sum_d \delta_{jk'}(t) (\mu_{ij}(t) / \lambda_{ij}(t)) \cdot \lambda_{ij}(t)} \]
\[ = \frac{\sum_d \delta_{jk}(t) \lambda_{ij}(t)}{\sum_{k'} \sum_d \delta_{jk'}(t) \lambda_{ij}(t)}, \quad \text{where } t = i + T_{ij}(i). \]

(17)

Note that the last equality in (17) is only valid when node \( i \) is unique as mentioned in assumption I that a single link \((i,j)\) is connected to the diverging node \( j \). The summation over \( i \) in the right-hand side is hence omitted here. And the \( \delta_{ij}(t) \) is the result of route choices and defined as
\[ \delta_{ij}^d(t) = \begin{cases} 1 & \text{if link}(i,j) \text{ lies on the shortest path from node } i \text{ to destination } d \text{ at time } t, \\ 0 & \text{otherwise}. \end{cases} \]

(18)

It is noticeable that diverging ratio \( \phi_{ij}(t) \) can be determined from the past information of \( \lambda_{ij}^d(t) \)'s once a path from node \( i \) to every destination \( \delta_{ij}^d(t) \) is found.

For a merging section at node \( j \), Eq. (15) is modified such that

\[ \mu_{jk}(t) = \operatorname{Min}[X_{ij}(t), \eta_{ij}(t) \cdot Y_{jk}(t)], \]

in which \( \eta_{ij}(t) \) (\( \leq 1 \)) is the merging capacity ratio. The \( \eta_{ij}(t) \) does not mean the ratio of merging flows at node \( j \) but stands for the share of downstream link capacity allocated to upstream link \( (i,j) \). The \( \eta_{ij}(t) \) may change over time since it depends not only the geometric design but also traffic demand flowing into the merging node \( j \). For example, the \( \eta_{ij}(t) \) for link \( (i,j) \) may increase if traffic demand on other merging links is light and vice versa.

When all merging links have queues (all merging links have sufficient demand), we may assume that \( \eta_{ij}(t) \) is equal to a static value of \( \eta_{ij}^* \) which is merging capacity ratio determined only from the geometric design of the merging section. However, when one or more merging links do not have queues, we may utilize the extra capacity to other links with sufficient demand as in (20). If the demand from other merging links, \( \sum_{r \neq i} X_{rj}(t) \), is not enough to fill out their share of \( (1 - \eta_{ij}^*) \cdot Y_{jk}(t) \), the remaining acceptable flow rate in \( Y_{jk}(t) \) can be utilized to link \( (i,j) \):

\[ \eta_{ij}(t) = \begin{cases} 1 - \frac{\sum_{r \neq i} X_{rj}(t)}{Y_{jk}(t)} & \text{if } \sum_{r \neq i} X_{rj}(t) < (1 - \eta_{ij}^*) \cdot Y_{jk}(t), \\ \eta_{ij}^* & \text{otherwise}. \end{cases} \]

(20)

This specification can be applicable to a situation where only one merging link has a queue. For more merging links with queues, you have to make an additional rule which assigns the extra capacity over two or more links with queues.

The above specifications of departure flow rates with physical queues could be arranged according to site specific characteristics of geometric design as well as flow condition. However, in general as shown, flow rates at diverging and merging sections would depend upon time-dependent flow conditions which are determined through the DUO assignment. In Section 4, we explain how these specifications are combined with the assignment.

4. Dynamic user optimal assignment

4.1. DUO principle

Every vehicle is assumed to choose the shortest route to its destination at any time based on the present link travel times. Let \( \pi_{id}(t) \) be the shortest travel time from node \( i \) to destination \( d \) prevailing at time \( t \), which means that \( \pi_{id}(t) \) is the sum of instantaneous link travel times, \( \tau_{ij}(t) \), along
the shortest route $p_{id}(t)$ evaluated at time $t$: $\pi_{id}(t) = \sum_{(i,j) \in p_{id}(t)} \tau_{ij}(t)$. Similar to the static assignment, the required condition for the DUO assignment is defined such that

$$
\begin{align*}
\pi_{id}(t) - \pi_{jd}(t) &= \tau_{ij}(t) & \text{if a vehicle with destination } d \text{ entering node } i \text{ at time } t \text{ uses link } (i,j), \\
\pi_{id}(t) - \pi_{jd}(t) &\leq \tau_{ij}(t) & \text{otherwise}.
\end{align*}
$$

(21)

This condition means that if a vehicle with destination $d$ entering node $i$ at time $t$ uses link $(i,j)$, node $j$ must be on the shortest route to destination $d$ evaluated from the instantaneous link travel times at present time $t$.

To eliminate the further complication, we here simply assume that every driver chooses a route based upon travel times. However, it is possible to introduce an appropriate cost of travel as a function of travel time and other relevant variables, but the core of the analysis remains as described in this paper.

According to the definition of the DUO assignment (21), the route choice of vehicles is clearly dependent only upon the instantaneous link travel times at present time $t$, but independent of the future link travel times. Therefore, the assignment is decomposed with respect to present time $t$; that is, the assignment can be considered sequentially from the beginning of the study time period.

Compared to the dynamic user equilibrium (DUE) assignment, the assignment in DUO is simple because prediction of future traffic condition is not required but users are assumed to choose their routes based on currently available information. And, one of the available information is instantaneous link travel time $\tau_{ij}(t)$ employed in this study as well as in many of previous researches.

4.2. Constructing cumulative curves with discrete time

A solution algorithm with point queues has been proposed using discrete times (Kuwahara and Akamatsu, 1997), but several modifications are required to deal with physical queues. In this section, the modified procedure to construct cumulative arrival and departure curves are summarized. The time axis is divided into small intervals of equal length $\Delta t$, and the arrival and departure rates of link $(i,j)$, $\lambda_{ij}(t)$ and $\mu_{ij}(t)$, are assumed constant during $[t, t + \Delta t]$.

Suppose that, at present time $t$, $\lambda_{ij}(t')$ and $\mu_{ij}(t')$ are assumed to be evaluated for every link $(i,j)$ and destination $d$ for $t' < t$. For example, as shown by thick lines in Fig. 4, the cumulative arrival and departure curves at every link are assumed to be obtained until $t' < t$. Let us now consider how to determine the arrival rate during $[t, t + \Delta t]$ and extend the cumulative curves $A_{ij}(t)$ until $t + \Delta t$ for $\forall(i,j)$ and $\forall d$ referring the proposed algorithm shown in Table 1.

In step 2, back-up-flag $\beta_{ij}(t)$, upstream demand rate $X_{ij}(t)$ and downstream acceptable flow rate $Y_{ij}(t)$ are evaluated from (12)–(14) for $\forall (i,j)$. From the known cumulative curves $A_{ij}(t)$ and $D_{ij}(t)$ by time $t$ and the flow-density relationship, back-up-flag $\beta_{ij}(t)$ can be first evaluated for every link. Then, using $\beta_{ij}(t)$, $X_{ij}(t)$ and $Y_{ij}(t)$ are determined.

In step 3, instantaneous link travel time $\tau_{ij}(t)$ is estimated from (11). Location of the shock $x_{ij}^{*}(t)$ is determined so as to satisfy (8) and then velocity on the link at every location $x$ in the congested region is estimated from (10).
Fig. 4. Construction of cumulative arrival and departure curves on link \((i,j)\).

Table 1

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Initialize link flow rates, cumulative curves, link travel times and present time</td>
</tr>
<tr>
<td>2</td>
<td>Determine back-up-flag (\beta_{ij}(t)), upstream demand rate (X_{ij}(t)) and downstream acceptable flow rate (Y_{ij}(t)) from (12)-(14) for (\forall (i,j))</td>
</tr>
<tr>
<td>3</td>
<td>Estimate instantaneous link travel time (\tau_{ij}(t)) from (11)</td>
</tr>
<tr>
<td>4</td>
<td>Find the shortest path from node (i) to destination (d) based on (\tau_{ij}(t)) and evaluate (\delta_{ij}^d(t)) for all ((i,j)) and (d) from the shortest route search</td>
</tr>
<tr>
<td>5</td>
<td>Determine diverging ratio (\phi_{ij}(t)) at diverging node (i) and merging capacity ratio (\eta_{ij}(t)) at merging node (j) from (17) and (20)</td>
</tr>
<tr>
<td>6</td>
<td>Determine departure flow rate (\mu_{ij}(t)) from (15), (16), and (19) considering queue backing up condition</td>
</tr>
<tr>
<td>7</td>
<td>Determine (\mu_{ij}^d(t)) for all ((i,j)) and (d) so as to satisfy the FIFO discipline (5)</td>
</tr>
</tbody>
</table>
| 8    | Determine the total arrival rate at node \(i\) for \([t,t+\Delta]\), \(\sum_j \lambda_{ij}^d(t)\), from the flow conservation (3):  
| 9    | Determine \(\lambda_{ij}^d(t)\) by loading \(\lambda_{ij}^d(t) = \lambda_{ij}^d(t) + \sum_i h_{ij}^d(t)\) onto a link starting from node \(i\) on the shortest path:  
| 10   | Extend \(A_{ij}^d(\cdot)\) and \(D_{ij}^d(\cdot)\) until time \(t+\Delta\) by straight lines with slopes \(\lambda_{ij}^d(t)\) and \(\mu_{ij}^d(t)\) |
| 11   | Update present time as \(t := t + \Delta\) and return to step 2 |
In step 4, using instantaneous link travel times $\tau_{ij}(t)$'s, the shortest route from every node $i$ to every destination $d$ is determined by a standard shortest route algorithm. From the shortest route search, we can determine $\delta^d_{ij}(t)$.

In step 5, diverging ratio $\phi_{ij}(t)$ at diverging node $i$ and merging capacity ratio $\eta_{ij}(t)$ at merging node $j$ are then determined from (17) and (20).

In step 6, departure flow rate $\mu_{ij}(t)$ is determined from (15), (16), and (19) considering queue backing up condition $\beta_{ij}(t)$.

In step 7, departure flow rate $\mu_{ij}(t)$ is allocated by destination to obtain $\mu^d_{ij}(t)$ so as to satisfy the FIFO discipline (5):

$$\mu^d_{ij}(t) = \mu_{ij}(t) \cdot \frac{\lambda^d_{ij}(t)}{\sum_d \lambda^d_{ij}(t)} , \quad t = \hat{t} + T_{ij}(\hat{t}), \quad \forall (i,j).$$

Since $\lambda^d_{ij}(t)$ and $D^d_{ij}(t)$ have been determined as piecewise linear lines at every $\Delta t$ interval until time $t$, $\hat{t}$, which may not be a multiple of $\Delta t$, can be determined based on the history of $\lambda^d_{ij}(t)$ and $D^d_{ij}(t)$ as shown in Fig. 4. This piecewise linear approximation, however, causes a bias in $\mu^d_{ij}(t)$, which in turn influences on values of instantaneous travel times and consequently on the route choice $\delta^d_{ij}(t)$. Since the bias could be accumulated over time and related to the route search, the analysis on the bias would not be easy and reserved for the future research. However, $\mu^d_{ij}(t)$'s are obtained so that the conservation is maintained: $\sum_d \mu^d_{ij}(t) = \mu_{ij}(t)$.

In step 8, the total arrival rate during $[t, t + \Delta t]$, $\sum_d \lambda^d_{ij}(t)$, is determined as $q_{id}(t) + \sum_h \mu^d_{hi}(t)$ from (3), since $q_{id}(t)$ and $\sum_h \mu^d_{hi}(t)$ have been already evaluated.

In step 9, $q_{id}(t) + \sum_h \mu^d_{hi}(t)$ is loaded onto a link on the shortest path from node $i$; that is, $\lambda^d_{ij}(t) = \delta^d_{ij}(t) \cdot [q_{id}(t) + \sum_h \mu^d_{hi}(t)]$, for all $(i,j)$ and $d$. If there are two or more equally shortest routes from node $i$ to $d$, the total rate of $q_{id}(t) + \sum_h \mu^d_{hi}(t)$ could be arbitrarily split onto any of the shortest routes. However, distribution of the total rate onto several equally shortest routes is not necessary to establish the DUO assignment principle defined above.

The demand of $q_{id}(t) + \sum_h \mu^d_{hi}(t)$ is loaded only onto the first downstream link $(i,j)$ from node $i$ on the shortest path during $[t, t + \Delta t]$ for all $(i,j)$ and $d$. Therefore, time interval $\Delta t$ should not be larger than travel time on link $(i,j)$ because otherwise the demand moves to the further downstream links. This is the reason why $\Delta t < \min_{(i,j)} c_{ij}/w_{ij}$ as shown in step 1.

In step 10, $A^d_{ij}(\cdot)$ and $D^d_{ij}(\cdot)$ for $\forall (i,j)$ and $\forall d$ can be extended from time $t$ to $t + \Delta t$ by straight lines with slopes $\lambda^d_{ij}(t)$ and $\mu^d_{ij}(t)$, respectively:

$$A^d_{ij}(t + \Delta t) = A^d_{ij}(t) + \lambda^d_{ij}(t) \cdot \Delta t , \quad A_{ij}(t + \Delta t) = \sum_d A^d_{ij}(t + \Delta t) ,$$

$$D^d_{ij}(t + \Delta t) = D^d_{ij}(t) + \mu^d_{ij}(t) \cdot \Delta t , \quad D_{ij}(t + \Delta t) = \sum_d D^d_{ij}(t + \Delta t) .$$

If the above procedure is repeated from the beginning of the study period, whole pictures of arrival and departure curves by destinations are drawn for all links.
5. An example

To easily examine the computation result and to demonstrate the difference between the physical and point queue analyses, let us consider a simple network with a freeway and an arterial running parallel as shown in Fig. 5. Suppose the evening commute problem where users are going back from the work places to residential area taking either the freeway or arterial. The characteristics of seven links are listed in Table 2 and the given OD demand is as follows:

- Origin 1 to destination 2: \[ q_{12}(t) = \begin{cases} 
 1000 \text{ (veh/h)} & 0 < t \leq 1 \text{ (h)} \text{ and } 3 < t \text{ (h)} \\
 2000 \text{ (veh/h)} & 1 < t \leq 3 \text{ (h)}
\end{cases} \]
- Origin 1 to destination 3: \[ q_{13}(t) = \begin{cases} 
 2000 \text{ (veh/h)} & 0 < t \leq 1 \text{ (h)} \text{ and } 3 < t \text{ (h)} \\
 4000 \text{ (veh/h)} & 1 < t \leq 3 \text{ (h)}
\end{cases} \]

Since the demand rate to destination 3 is at most 4000 (veh/h) but \(f_{63}^{\text{max}}\) of links (6, 3) is limited to 3000 (veh/h), a queue likely forms from node 6 on freeway link (5, 6). And if the queue reaches upstream link (4, 5), departure flow rate \(\mu_{45}(t)\) would be affected by the physical queue; that is, the functional form of \(\mu_{45}(t)\) is given by (16) for the diverging section.

Figs. 6 and 7 illustrate the cumulative curves on links for demand from 1 to 3 and from 1 to 2 separately with time interval \(\Delta t = 0.01\) (h). Since free flow travel time via the freeway is faster than via the arterial (to destination 3, 0.4 and 0.7 (h); and to destination 2, 0.3 and 0.4 (h), respectively), everyone uses the freeway at the beginning. At time 1.35, a queue forms from node 6 because the demand rate of 4000 (veh/h) to destination 3 exceeds the departure flow rate \(\mu_{56}(t)\):

![Fig. 5. An example network. figures in ( ) = \(\ell_{ij}/w_{ij}\) (h).](image)

<table>
<thead>
<tr>
<th>Link</th>
<th>Length (km)</th>
<th>(\ell_{ij}/w_{ij}) (h)</th>
<th>(\ell_{ij}/w_{ij}') (h)</th>
<th>(k_{ij}^{\text{max}}) (veh/km)</th>
<th>(f_{ij}^{\text{max}}) (veh/h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 4)</td>
<td>2.0</td>
<td>0.05</td>
<td>0.1</td>
<td>450.0</td>
<td>6000.0</td>
</tr>
<tr>
<td>(4, 5)</td>
<td>16.0</td>
<td>0.2</td>
<td>0.8</td>
<td>375.0</td>
<td>6000.0</td>
</tr>
<tr>
<td>(5, 6)</td>
<td>8.0</td>
<td>0.1</td>
<td>0.4</td>
<td>250.0</td>
<td>4000.0</td>
</tr>
<tr>
<td>(6, 3)</td>
<td>2.0</td>
<td>0.05</td>
<td>0.1</td>
<td>225.0</td>
<td>3000.0</td>
</tr>
<tr>
<td>(1, 2)</td>
<td>16.0</td>
<td>0.4</td>
<td>0.8</td>
<td>450.0</td>
<td>6000.0</td>
</tr>
<tr>
<td>(2, 3)</td>
<td>12.0</td>
<td>0.3</td>
<td>0.6</td>
<td>450.0</td>
<td>6000.0</td>
</tr>
<tr>
<td>(5, 2)</td>
<td>2.0</td>
<td>0.05</td>
<td>0.1</td>
<td>450.0</td>
<td>6000.0</td>
</tr>
</tbody>
</table>
\[ \mu_{56}(t) = \text{Min}[X_{56}(t), Y_{63}(t)] = \text{Min}[4000, 3000] = 3000 \text{ (veh/h)}. \]

The shock reaches node 5 at time 1.75 \((= 1.35 - \frac{\ell_{56}}{w_{56}'} = 1.35 + 8/20)\). Since link \((5, 6)\) has been used at its capacity, arrival time of the backward wave just happens to be equal to time when the shock occurs at node 5. By this time, instantaneous link travel time \(\tau_{56}(t)\) becomes to 0.26 (h) but no longer increases because the physical queue fully occupies the link. And since still 0.26 (h) is less than \(\tau_{33}(t) = 0.3 \text{ (h)}\), every one to destination 3 takes links \((5, 6)\) and \((6, 3)\) instead of links \((5, 2)\) and \((2, 3)\). Indeed, off-ramp link \((5, 2)\) is never used by demand for destination 3 throughout the study period as shown in Fig. 6. Therefore, from (17), the diverging ratios at node 5 still remain as the following values:

\[ \phi_{56}(t) = \frac{z_{45}^3(\hat{t})}{z_{45}^3(\hat{t}) + z_{45}^3(\hat{t})} = \frac{4000}{2000 + 4000} = 2/3, \]

\[ \phi_{52}(t) = \frac{z_{45}^2(\hat{t})}{z_{45}^2(\hat{t}) + z_{45}^3(\hat{t})} = \frac{2000}{2000 + 4000} = 1/3, \quad \text{where } \hat{t} = \hat{t} + T_{45}(\hat{t}). \]
Cumulative Trips

Fig. 7. Cumulative arrivals and departures of demand for destination 2 (physical queues, Δt = 0.01 (h)).

However, from this time, the departure flow rate of link (4, 5), \(μ_{45}(t)\), drops according to (16):

\[
X_{45}(t) = f_{45}^{\text{max}},
\]

\[
Y_{56}(t) = \{1 - \beta_{56}(t)\} \cdot f_{56}^{\text{max}} + \beta_{56}(t) \cdot μ_{56}(t - \ell_{56}/w_{56}) = \mu_{56}(t - \ell_{56}/w_{56}),
\]

\[
Y_{52}(t) = \{1 - \beta_{52}(t)\} \cdot f_{52}^{\text{max}} + \beta_{52}(t) \cdot μ_{52}(t - \ell_{52}/w_{52}) = f_{52}^{\text{max}}.
\]

\[
μ_{45}(t) = \text{Min} \left[ X_{45}(t), \sum_k Y_{5k}(t), \text{ Min}_k \{Y_{5k}(t)/φ_{5k}(t)\} \right]
\]

\[
= \text{Min}[f_{45}^{\text{max}}, \mu_{56}(t - \ell_{56}/w_{56}) + f_{56}^{\text{max}}, \mu_{56}(t - \ell_{56}/w_{56})/φ_{56}(t), f_{52}^{\text{max}}/φ_{52}(t)]
\]

\[
= \text{Min}[6000, 3000 + 6000, 3000/(2/3), 6000/(1/3)] = 4500 \text{ (veh/h)}.
\]

The arrival rates to links (5, 6) and (5, 2) are hence restricted and the departure rate for destination 3 drops from 4000 to 3000 and for destination 2 from 2000 to 1500 as shown in the figures.

However, everyone access to node 5, the queue on link (4, 5) keeps growing and link travel time \(τ_{45}(t)\) reaches 0.3 (h) at time 2.01. Then, vehicles to destination 2 shift from the freeway to the arterial and hence \(A_{42}^2(t)\) starts increasing as shown in Fig. 7. However, \(τ_{45}(t)\) keeps increasing, since the arrival rate for destination 3 at link (4, 5) is still 4000 and the corresponding departure rate remains at 3000 even after time 2.01. The departure flow rate from link (4, 5) is always 3000 for demand 1 to 3, independent of route choice by demand 1 to 2. \(T_{45}(t)\) is indeed dictated by demand from 1 to 3.

At time 2.09, \(τ_{45}(t)\) becomes 0.34. Demand to 3 then switches the route to the arterial, since \(τ_{56}(t)\) stays in 0.26 after time 1.75. Since some amount of demand has already chosen the freeway
route by this time, $\tau_{45}(t)$ keeps increasing and the entire demand thus uses the arterial for a while. This time lag in route change is normally observed in the DUO assignment in contrast with the DUE assignment.

Due to the route shift of demand 1 to 3, $\tau_{45}(t)$ turns to decreasing and comes back to 0.34 at time 2.29, and demand to 3 returns to the freeway. Then, $\tau_{45}(t)$ becomes 0.3 at time 2.34 and demand 1 to 2 switches to freeway also. However, the travel time is still decreasing due to the time lag in DUO. At some time later, $\tau_{45}(t)$ turns to increasing again and when $\tau_{45}(t)$ becomes 0.3, demand to 2 switches to the arterial. But again, the travel time keep increasing due to the lag, and when $\tau_{45}(t)$ returns to 0.34, demand to 3 switches to the arterial. This route switching is observed twice until the demand decreases at time 3.0.

Fig. 8 shows the same physical queue evolution but using time interval $\Delta t = 0.04$ (h). Although minor differences from Fig. 6 are observed especially route switching timings, the shapes of cumulative curves and general sequences of route choices seem to be reproduced even with $\Delta t = 0.04$ (h).

On the other hand, Fig. 9 illustrates the cumulative curves for demand 1 to 3 under the point queue concept. In this situation, as shown in Fig. 1, a queue vertically forms at the end of link (5, 6) and only $\tau_{56}(t)$ increases but the queue never expands to other links. Therefore, for destination 2, everyone always takes the freeway and the arterial link (1, 2) is never used. For destination 3, although there are three alternative routes: route $A =$ links (1, 4), (4, 5), (5, 6), (6, 3),

![Cumulative Trips](image)

Fig. 8. Cumulative arrivals and departures of demand for destination 3 (physical queues, $\Delta t = 0.04$ (h)).
route B = links (1, 4), (4, 5), (5, 2), (2, 3) and route C = links (1, 2), (2, 3), the arterial route C is never used from the same reason. (In contrast, with physical queues, route B is never used because \( \tau_{56}(t) \) does not get longer than 0.26 due to the fully occupied physical queue on the link.) Since everyone to destination 3 first takes route A, a queue developed on link (5, 6) is gradually getting longer and route A travel time becomes equal to one on route B at time 1.85. Then, very frequent route switching starts so as to equalize the travel times until the demand decreases after time 3.0.

This simple example illustrates a typical effect of physical queues compared to point queues. Especially, with physical queues, active alternative routes could be different from those under the point queue concept. In this example, once freeway link (5, 6) is fully occupied by a queue, its travel time no longer increases and more traffic tends to use freeway than under the point queue concept. The similar effect has been reported by Daganzo (1998). Another interesting effect with physical queues is that delay could be imposed on vehicles which do not pass a bottleneck. Here, vehicles to destination 2 are also caught by a queue on link (4, 5) and delayed, while they are never caught by a queue under the point queue concept.

The DUO assignment can be considered as a simplified representation of route choices brought by ATIS (Advanced Traveler Information Systems) in ITS, where variable message signs, highway radios, in-vehicle navigation equipment, etc. frequently supply current traffic information. However, we recognize that there is still a gap between DUO and the dynamic system optimum (DSO) situations. Generally speaking, for a network with a freeway and an arterial running parallel as in this example, the basic policy for the system optimum is to assign demand onto the
freeway just equal to its capacity but not more than the capacity. The reason is that once a queue forms due to the excess demand of even a little at one time, the queue lasts long even if the demand rate is maintained at its capacity value thereafter. We may need better control strategies rather than just to provide current traffic information by ATIS.

6. Summary and future research needs

This paper extends the DUO assignment with point queues to one with physical queues so as to describe queue backing up phenomena. Given time dependent many-to-many OD volumes, we first formulate required constraints of the flow conservation and the first-in-first-out queue discipline. Then, based on the kinematic wave theory, the physical queue evolution is included into the DUO assignment. Finally, a solution algorithm is proposed and applied to a simple network to demonstrate the difference between physical and point queue analyses.

For the future research needs, it would be needed to include the physical queue analysis to the dynamic user equilibrium assignment, in which users choose routes based upon actually experienced route costs. The DUE assignment cannot be completed only past and present traffic information and the problem is thus no longer decomposed by current time, but we have to consider the whole study period. For DUE under the point queue concept with a one-to-many (or a many-to-one) OD pattern, Kuwahara and Akamatsu (1993, 1994) proposed the formulation and a solution algorithm by decomposing the problem with respect to starting time from a single origin (or arrival time at a single destination). The inclusion of the physical queues into this kind of the DUE assignment with a one-to-many OD pattern would be an immediate next topic. Also, the DUE assignment so as to handle a many-to-many OD under the point or physical queue concept would be interesting future topic.

Acknowledgements

The authors would like to express deep appreciation to anonymous referees for their valuable and constructive comments on the earlier versions of this paper.

References

