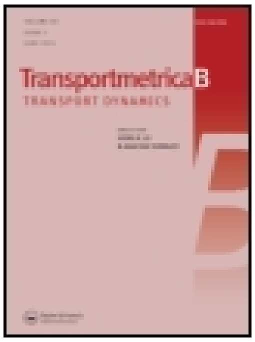


Transportmetrica B: Transport Dynamics

ISSN: 2168-0566 (Print) 2168-0582 (Online) Journal homepage: http://www.tandfonline.com/loi/ttrb20

Multiclass, simultaneous route and departure time choice dynamic traffic assignment with an embedded spatial queuing model

Xianyuan Zhan & Satish V. Ukkusuri

TRANSPORTMETRICA B: TRANSPORT DYNAMICS, 2017
https://doi.org/10.1080/21680566.2017.1354738

Multiclass, simultaneous route and departure time choice dynamic traffic assignment with an embedded spatial queuing model

Xianyuan Zhan and Satish V. Ukkusuri

Lyles School of Civil Engineering, Purdue University, West Lafayette, IN, USA

ABSTRACT
This paper develops a complementarity formulation for a multi-user class, simultaneous route and departure time choice dynamic user equilibrium (DUE) model. A path-based multiclass cell transmission model (mCTM) is embedded to propagate the traffic flow on the network. Heterogeneous user classes are incorporated in the new formulation and heterogeneity is based on different preferred arrival times, cost perception for travel time, early and late arrival penalties. Multiple model properties have been showed. The proposed model is solved as an equivalent non-monotone variational inequality (VI) problem defined on a product set. A modified proximal point algorithm is used to solve the proposed non-monotone VI problem. Numerical results show that the solution approach is able to find the equilibrium or close to equilibrium solutions. The new formulation and solution approach show the feasibility of solving the multiclass DUE problem for general traffic networks.

ARTICLE HISTORY
Received 23 August 2016
Accepted 9 July 2017

KEYWORDS
Dynamic traffic assignment; transportation networks; proximal point algorithm; multi-user class; spatial queue

1. Introduction

The importance of incorporating multiple user classes in transportation network modeling has long been realized. The fact that transportation systems have many classes of users has motivated many researchers to develop models and integrated frameworks that capture heterogeneous user behaviors. Dafermos (1972) was among the first researchers to propose the notion of multiclass traffic equilibrium. The notion states that all classes of users are to be assigned to the network so that no one in each class can improve his or her travel cost by unilaterally changing his or her route. This is a natural extension of Wardrop's first principle from single user class to multiple user classes. While there are comprehensive studies on static multiclass traffic assignment, the literature on dynamic multiclass traffic assignment is limited.

Compared with static traffic assignment, dynamic traffic assignment (DTA) shows great advantage by capturing more realistic traffic flow propagation characteristics and incorporating time dimension into the analysis. DTA models have seen substantial development since the pioneering work of Merchant and Nemhauser (1978). After 30 years of development, the DTA models in the literature can be classified into four broad methodological groups as suggested by Peeta and Ziliaskopoulos (2001): mathematical programming, optimal control, variational inequality (VI) and simulation-based approach. Under such context, studying the multiclass dynamic user equilibrium (DUE) is particularly useful in analyzing the travel behaviors and complex interactions of heterogeneous users in the traffic network. This has important application in morning commuting problem, in which multiple

CONTACT Satish V. Ukkusuri sukkusur@purdue.edu

user groups with different values of time (VOT) and preferred arrival times traveling on a congested network.

In the literature, various multiclass DTA models with different definitions of user classes have been developed to incorporate the interactions among heterogeneous users in the traffic network. Based on the differences on the behavioral assumptions on user classes, these multiclass DTA models can be classified into three categories: (1) users with different vehicle characteristics, e.g. cars and trucks (Bliemer and Bovy 2003; Bliemer 2007; Mesa-Arango and Ukkusuri 2014); (2) users with a different route choice behavior (Lo, Ran, and Hongola 1996; Ran, Lee, and Shin 2002; Szeto, Jiang, and Sumalee 2011) and (3) users with different preferred arrival times and VOT (Ramadurai et al. 2010; Han, Ukkusuri, and Doan 2011; Liu and Nie 2011; Pang et al. 2012).

Lo, Ran, and Hongola (1996) and Ran, Lee, and Shin (2002) were among the first to formulate the multiclass DTA model as a mathematical programming problem through a VI approach. Travelers are classified solely based on their route choice behavior, who follows: (a) predetermined or fixed routes, (b) stochastic dynamic user-optimal assignment and (c) dynamic user-optimal assignment. Strong user behavioral assumptions were made in this model, which lacks flexibility in modeling realistic travel patterns in the network. From a different perspective, Bliemer and Bovy (2003) proposed a quasi-VI DTA model to capture the interactions between user groups with different vehicle characteristics, e.g. cars and trucks. Different link travel time functions were used for each user class, and only route choice was considered in this model. Bliemer (2007) proposed a new analytical multiclass dynamic network loading model as part of a simulation-based DTA model, in which the dynamic queuing, spillback and the heterogeneity in vehicle characteristics are captured.

Ramadurai et al. (2010) developed a single bottleneck model with heterogeneous commuters as a linear complementarity problem (CP) based on Vickrey's (1969) single bottleneck model. Liu and Nie (2011) studied the morning commute problem with heterogeneous users using the single bottleneck model with two parallel paths. Analytical solutions of no-toll equilibrium, system optimal (SO) and time-based SO were examined. Pang et al. (2012) introduced a linear complementarity system formulation for a continuous-time multiclass single bottleneck DUE model. Analysis on solution existence and numerical solution scheme were discussed in the paper. The single bottleneck model is a convenient and tractable tool to analyze the heterogeneous commuters' travel behavior on a single link. However, the single bottleneck assumption limits its application in general traffic networks.

The introduction of cell transmission model (CTM) (Daganzo 1994, 1995) has provided a new opportunity and framework to account for the spatial queuing behavior in DTA modeling (Lo and Szeto 2002; Szeto and Lo 2004; Han, Ukkusuri, and Doan 2011; Szeto, Jiang, and Sumalee 2011; Doan and Ukkusuri 2012; Ukkusuri, Han, and Doan 2012). Several efforts have been made to incorporate multiclass analysis in the cell-based DTA framework. Tuerprasert and Aswakul (2010) proposed a multiclass cell transmission model (mCTM) for heterogeneous users in the traffic network. However, the mCTM was proposed only as a simulation-based network loading model, without being applied to any traffic assignment model. Szeto, Jiang, and Sumalee (2011) proposed a cell-based multiclass DTA problem, which was formulated as a fixed-point problem. The model embeds a Monte Carlo-based stochastic CTM to capture the effect of physical queues and the random evolution of traffic states during flow propagation. The user classes are classified based on different levels of perception error on travel time. Still, only route choice is modeled and the departure time choice is not considered. Han, Ukkusuri, and Doan (2011) proposed a DUE problem with an embedded CTM. The model is capable of capturing departure time choice, elastic demand as well as user heterogeneity. The weakness lies in that it is only applicable to networks with single origin–destination (OD) and multiple parallel paths. Ukkusuri, Han, and Doan (2012) extended Han's work to general traffic networks. A complementarity formulation for the CTM-based DUE model with both departure time and route choices was proposed. However, due to the increase of modeling complexity in the formulation, only homogeneous users are considered.

Motivated from previous studies, we propose a CP formulation for a multi-user class, simultaneous route and departure time choice DUE model. The heterogeneity setting among multi-user classes follows the works of Cohen (1987), Arnott, de Palma, and Lindsey (1988, 1994), and Szeto and Lo (2004) that focuses on user groups with different preferred arrival times, cost perception for travel time, early and late arrival penalties. Different vehicle types are not considered in this study. By incorporating both the route choice and departure time choice in the model, detailed travel behaviors for heterogeneous users under equilibrium condition can be obtained. A path-based mCTM is developed and embedded to propagate the traffic flow on the network. The spatial queuing on links and traffic spillback are captured. The model is solved as an equivalent VI problem. This model is built on works of Han, Ukkusuri, and Doan (2011) and Ukkusuri, Han, and Doan (2012). While the previous models only subject to restricted network or single user class, the proposed formulation extends the applicability to multiuser class and general networks. Various model properties, such as OD level first-in-first-out (FIFO), continuity, non-monotonicity and solution existence, have been proved. A modified proximal point algorithm (PPA) is used to solve the non-monotone, non-differentiable VI problem, and has shown to converge to the exact or close to equilibrium solutions on the test networks.

This work contributes to the literature in following aspects:

(1) We propose a new multiclass CTM (mCTM) for general networks, which allows for tracking detailed sub-flow from each user group in the flow propagation.
(2) We develop a new multiclass DUE model that is formulated as a CP and prove the solution existence along with many other model properties.
(3) We highlight the non-monotonicity and discontinuity issues of the average travel time function. Such issues are common in many CTM-based DTA models, but often overlooked or not well addressed. We also propose a simple background flow approach to handle the discontinuity issue.
(4) We develop a modified PPA to solve the large-scale non-monotone VI problem. Extensive numerical experiments are conducted on multiple networks and different demand settings to demonstrate the effectiveness of the proposed algorithm.

The rest of the paper is organized as follows. Section 2 introduces the formulations of mCTM and the CP formulations of the DUE model. Section 3 shows various model properties and Section 4 further proves the solution existence. Section 5 presents the solution approach of the proposed problem. Multiple numerical experiments are conducted and analyzed in Section 6 and Section 7 concludes the paper.

2. Problem formulation

2.1. Model description

We consider a DUE problem in a general network with multiple OD pairs and heterogeneous users. For each OD pair and user group, a number of 'selfish' drivers travel from origin to destination by selecting the route and departure time with the least cost. The travel cost is composed of three parts: travel time cost, early and late penalties. Higher late penalty is considered compared with early arrival penalty. The user groups are differentiated by preferred arrival times, cost perception for travel time, early and late arrival penalties rather than vehicle types. The mCTM formulations and the CP formulation of the DUE problem are presented as follows.

2.2. Notations

Indices:

w index for origin–destination pairs
p index for paths
g index for user groups
i, j index for cells
t index for time interval

Parameters:

$\alpha_g^w, \beta_g^w, \gamma_g^w$	unit cost of travel time, early and late arrival for OD pair w and user group g. $\beta_g^w < \alpha_g^w < \gamma_g^w$
t_g^{*w}	preferred arrival time for OD pair w and user group g
d_g^w	total demand from OD pair w and user group g
η	infinitesimal parameter to avoid zero denominator
T, T_f	maximum departure time and overall time horizon
N^i	jam density of cell i
Q^i	flow capacity out of cell i
δ	Backward-to-forward shockwave propagation ratio

Sets:

C	set of cells. $C_{\mathrm{O}}, C_{\mathrm{R}}, C_{\mathrm{S}}, C_{\mathrm{D}}, C_{\mathrm{M}}$ denote sets of ordinary, source, sink, diverging, merging cells
E	set of links. $E_{\mathrm{O}}, E_{\mathrm{D}}, E_{\mathrm{M}}$: sets of ordinary, diverging, merging links
G	set of user groups
Γ_i^{-1}	set of predecessors of cell i
Γ_i	set of successors of cell i
W	set of all OD pairs
P^w	set of paths from OD pair w
P	set of all paths, $P = \cup_{w \in W} P^w$
$\mathbb{T}$	set of all time intervals up to T, $\mathbb{T} \triangleq \{0, \ldots, T\}$
$\mathbb{T}_f$	set of all time intervals up to T_f, $\mathbb{T}_f \triangleq \{0, \ldots, T_f\}$

Variables:

$x_{p,g,t}^i$	cell occupancy of cell i at time t of user group g for the flow on path p
$y_{p,g,t}^{i,j}$	flow from cell i to j time t of user group g for the flow on path p
$\bar{x}_t^i$	aggregate cell occupancy of cell i at time t. $\bar{x}_t^i = \sum_g \sum_p x_{p,g,t}^i, \forall i \in C, t \in \mathbb{T}_f$
$\bar{y}_t^{i,j}$	aggregate flow from cell i to j at time t. $\bar{y}_t^{i,j} = \sum_g \sum_p y_{p,g,t}^{i,j}, \forall (i,j) \in E, t \in \mathbb{T}_f$
$\tilde{x}_t^{i,j}$	aggregate cell occupancy from cell i to j at time t going to cell j
$r_{p.g.t}$	departure rate at time t for the flow using path p
$TT_{p,t}^a$	average travel time for the flow using path p at time t
$\upsilon_{p,t,t'}$	auxiliary variable for average travel time estimation

2.3. Multiclass path-based cell transmission model (mCTM)

The multiclass path-based CTM model developed in this work is a multiclass extension of the single user group CTM formulation in Ukkusuri, Han, and Doan (2012). In this model, each link is divided into homogeneous cells with the length equal to the distance traveled at free-flow speed within a single simulation interval. The amount of flow inside a cell and the flow from one cell to another cell are modeled as cell occupancy $x_{p,g,t}^i$ and flow $y_{p,g,t}^{i,j}$ with index on paths, user groups and time steps. By tracking the flow of each user group, the multiclass flow is properly propagated within the network. The detailed formulations of mCTM are presented as follows:
Initialization

$$x_{p,g,0}^i = 0, \quad y_{p,g,0}^{i,j} = 0, \quad \forall i \in C, (i,j) \in E, p \in P, g \in G \tag{1}$$

Source cell C_R

$$x_{p,g,t}^i = r_{p,g,t} + x_{p,g,t-1}^i - y_{p,g,t-1}^{i,j}, \quad \forall i \in C_{\mathrm{R}}, i \in p, g \in G, j \in \Gamma_i, \ t = \{1, \ldots, T+1\} \tag{2}$$

$$x^i_{p,g,t} = x^i_{p,g,t-1} - y^{i,j}_{p,g,t-1}, \quad \forall i \in C_O, i \in p, g \in G, j \in \Gamma_i,\ t = \{T+2, \ldots, T_f\} \tag{3}$$

Also, the demand for each OD pair satisfies:

$$\sum_{p \in P^w} \sum_{t=0}^{T} r_{p,g,t} = d^w_g, \quad \forall w \in W, g \in G \tag{4}$$

Ordinary cells C_O

The ordinary cells are cells with single predecessor and successor. The cell occupancy is updated as follows:

$$x^i_{p,g,t} = x^i_{p,g,t-1} + y^{k,i}_{p,g,t-1} - y^{i,j}_{p,g,t-1}, \quad \forall i \in C_O, i \in p, g \in G, k \in \Gamma^{-1}_i, j \in \Gamma_i, t \in \mathbb{T}_f \backslash \{0\} \tag{5}$$

Diverging/merging cells C_D, C_M and C_{DM}

We define cells with single predecessor and multiple successor as diverging cells C_D; cells with multiple predecessor and single successor as merging cell C_M; cells with multiple predecessors and successors as diverging/merging cells C_{DM}. For diverging and merging cells ($i \in C_D \cup C_M \cup C_{DM}$), the following equation applies:

$$x^i_{p,g,t} = x^i_{p,g,t-1} + \sum_{k \in \Gamma^{-1}_i} y^{k,i}_{p,g,t-1} - \sum_{j \in \Gamma_i} y^{i,j}_{p,g,t-1}, \quad \forall \{k,i,j\} \subset p; g \in G, t \in \mathbb{T}_f \backslash \{0\} \tag{6}$$

Sink cells C_S

$$x^i_{p,g,t} = x^i_{p,g,t-1} + y^{k,i}_{p,g,t-1}, \quad \forall i \in C_S, i \in p; g \in G, k \in \Gamma^{-1}_i, t \in \mathbb{T}_f \backslash \{0\} \tag{7}$$

Ordinary links E_O

The aggregate flow $\bar{y}^{i,j}_t$ satisfies:

$$\bar{y}^{i,j}_t = \min\,(\bar{x}^i_t, Q^i, Q^j, \delta(N^j - \bar{x}^j_t)), \quad \forall (i,j) \in E_O, t \in \mathbb{T}_f \backslash \{0\} \tag{8}$$

where $\bar{x}^i_t$ is the aggregated cell occupancy of cell i at time t, computed as

$$\bar{x}^i_t = \sum_{g \in G} \sum_{\forall p \ni i} x^i_{p,g,t} \tag{9}$$

At the disaggregate level, the flow can be represented as

$$y^{i,j}_{p,g,t} = \min\,(\bar{x}^i_t, Q^i, Q^j, \delta(N^j - \bar{x}^j_t)) \times \frac{x^i_{p,g,t}}{\bar{x}^i_t + \xi}, \quad \forall (i,j) \in E_O, p \ni i, g \in G, j \in \Gamma_i, t \in \mathbb{T}_f \backslash \{0\} \tag{10}$$

where ξ is an infinitesimal positive number to avoid dividing by 0 error.

Diverging links E_D

For the flow on diverging links, we follow the strategy discussed in Ukkusuri, Han, and Doan (2012), in which the path flow $y^{i,j}_{p,g,t}$ is determined using the proportion of path-based cell occupancy $x^i_{p,g,t}$ in the aggregate cell occupancy $\tilde{x}^{i,j}_t$ at cell i and time t oriented to cell j

$$\tilde{x}^{i,j}_t = \sum_{g \in G} \sum_{\forall p \ni (i,j)} x^i_{p,g,t}, \quad \forall i \in C_D, j \in \Gamma_i, t \in \mathbb{T}_f \backslash \{0\} \tag{11}$$

$$\bar{y}^{i,j}_t = \min\,(\tilde{x}^{i,j}_t, Q^j, \delta(N^j - \bar{x}^j_t)) \times \min\left(1, \frac{Q^i}{\sum_{j' \in \Gamma_i} (\min(\tilde{x}^{i,j'}_t, Q^{j'}, \delta(N^{j'} - \bar{x}^{j'}_t))) + \xi}\right) \tag{12}$$

$$y^{i,j}_{p,g,t} = \bar{y}^{i,j}_t \times \frac{x^i_{p,g,t}}{\tilde{x}^{i,j}_t + \xi}, \quad \forall i \in C_D, p \ni i, g \in G, j \in \Gamma_i, t \in \mathbb{T}_f \backslash \{0\} \tag{13}$$

Merging links E_M

Similar to the diverging case, the flow for the merging links E_M can be written as

$$\bar{y}_t^{k,i} = \min(Q^k, \bar{x}_t^k) \times \min\left(1, \frac{\min(Q^i, \delta(N^i - \bar{x}_t^i))}{\sum_{k' \in \Gamma_i^{-1}} (\min(Q^{k'}, \bar{x}_t^{k'})) + \xi}\right) \tag{14}$$

$$y_{p,g,t}^{k,i} = \bar{y}_t^{k,i} \times \frac{x_{p,g,t}^k}{\bar{x}_t^k + \xi}, \quad \forall i \in C_M, p \ni i, g \in G, k \in \Gamma_i^{-1}, t \in \mathbb{T}_f \backslash \{0\} \tag{15}$$

2.4. Travel time estimation

In DTA problems, the path travel times are typically used as the delay operator to measure the cost of the flow departing at any given time steps. This work follows the idea of computing average path travel time from the cumulative departure and arrival pattern that discussed in Han, Ukkusuri, and Doan (2011) and Ukkusuri, Han, and Doan (2012). A similar approach is also used in Lo and Szeto (2002) and Szeto and Lo (2004). However, we make several modifications to improve the accuracy as well as guarantee the continuity of the average travel time computation. The average path travel time is computed using the following equations:

$$\upsilon_{p,t,t'} = \max\left(0, \sum_{g \in G} \sum_{h=0}^{t} (r_{p,g,h} + \mu) - \sum_{g \in G} x_{p,g,t'}^s\right), \quad \forall t' = t, \ldots, T_f \tag{16}$$

$$TT_{p,0}^a = \frac{\sum_{h=0}^{T_f-1} (\upsilon_{p,0,h} - \upsilon_{p,0,h+1})(h+1)}{\sum_{g \in G} (r_{p,g,0} + \mu)}, \quad \forall p \in P \tag{17}$$

$$TT_{p,t}^a = \frac{\sum_{h=t}^{T_f-1} (\upsilon_{p,t,h} - \upsilon_{p,t,h+1} + \upsilon_{p,t-1,h+1} - \upsilon_{p,t-1,h})(h+1-t)}{\sum_{g \in G} (r_{p,g,t} + \mu)}, \quad \forall p \in P, t \in \mathbb{T} \tag{18}$$

The variable $\upsilon_{p,t,t'}$ is the amount of flow that departs before time t but has not arrived at the sink cell s of path p at time t'. The term $\upsilon_{p,t,h} - \upsilon_{p,t,h+1} + \upsilon_{p,t-1,h+1} - \upsilon_{p,t-1,h}$ corresponds to the amount of the flow departed at time t and arrived at time step $h+1$, which has path travel time equal to $h - t + 1$. Hence, the computation of $TT_{p,t}^a$ averages the path travel time using the weight of the proportion of flow arrived. Furthermore, we introduce an independent fixed background flow of amount μ assigned on each path and user group in addition to the investigated departure flow $r_{p,g,t}$. Different from Lo and Szeto (2002), Szeto and Lo (2004), Han, Ukkusuri, and Doan (2011) and Ukkusuri, Han, and Doan (2012), we take μ to be a small positive value, but do not allow μ to take an arbitrarily small value ($\mu \not\to 0$). As we will show in the later section, the value μ plays an important role on the continuity and the smoothness of the average path travel time function $TT_{p,t}^a(\boldsymbol{r})$.

One can notice that the group index is not appearing in $TT_{p,t}^a$ calculation. This is because different user groups are homogeneous in terms of vehicle characteristics and travel time computation. Thus, different groups of users on the same path departing at the same time will experience the same travel time. The equivalent complementarity condition of Equation (16) can be written as

$$0 \leq \upsilon_{p,t,t'} \perp \upsilon_{p,t,t'} - \left(\sum_{g \in G} \sum_{h=0}^{t} (r_{p,g,h} + \mu) - \sum_{g \in G} x_{p,g,t'}^s\right) \geq 0, \quad \forall p \in P, t \in \mathbb{T}, t' = t, \ldots, T_f \tag{19}$$

2.5. Overall model

2.5.1. Dynamic equilibrium condition

In analogy to the static multiclass user equilibrium, the multiclass DUE is defined as: no flow in each user group of each OD pair can improve its travel cost by unilaterally changing its route and departure time. The complementarity formulation of the equivalent multiclass DUE condition is defined as:

$$0 \le r_{p,g,t} \perp \alpha_g^w TT_{p,t}^a + \beta_g^w e_{p,g,t} + \gamma_g^w l_{p,g,t} - C_g^{*w} \ge 0, \quad \forall w \in W, g \in G, p \in P^w, t \in \mathbb{T} \tag{20}$$

where α_g^w, β_g^w and γ_g^w are unit cost for average path travel time, early and late arrival penalties for user group g and OD pair w, which satisfy $\beta_g^w < \alpha_g^w < \gamma_g^w$. $e_{p,g,t}, l_{p,g,t}$ are early and late arrival times, computed as

$$e_{p,g,t} = \max(0, t_g^{*w} - t - TT_{p,t}^a), \quad \forall w \in W, p \in P^w, g \in G, t \in \mathbb{T} \tag{21}$$

$$l_{p,g,t} = \max(0, t + TT_{p,t}^a - t_g^{*w}) = e_{p,g,t} - (t_g^{*w} - t - TT_{p,t}^a), \quad \forall w \in W, g \in G, p \in P^w, t \in \mathbb{T} \tag{22}$$

Equation (21) can be converted into the following complementarity condition:

$$0 \le e_{p,g,t} \perp e_{p,g,t} - (t_g^{*w} - t - TT_{p,t}^a) \ge 0, \quad \forall w \in W, g \in G, p \in P^w, t \in \mathbb{T} \tag{23}$$

Furthermore, the demand satisfaction condition must be satisfied under the equilibrium,

$$0 \le C_g^{*w} \perp \sum_{p \in P^w} \sum_{t=0}^{T} r_{p,g,t} - d_g^w \ge 0, \quad \forall w \in W, g \in G \tag{24}$$

2.5.2. Overall formulation for the DUE model

Summarizing all the formulations discussed above, the overall formulation of the problem is listed below:

$$\begin{aligned}
&0 \le r_{p,g,t} \perp \alpha_g^w TT_{p,t}^a + \beta_g^w e_{p,g,t} + \gamma_g^w l_{p,g,t} - C_g^{*w} \ge 0 \\
&0 \le e_{p,g,t} \perp e_{p,g,t} - (t_g^{*w} - t - TT_{p,t}^a) \ge 0 \\
&0 \le C_g^{*w} \perp \sum_{p \in P^w} \sum_{t=0}^{T} r_{p,g,t} - d_g^w \ge 0 \\
&0 \le \upsilon_{p,t,t'} \perp \upsilon_{p,t,t'} - \left(\sum_{g \in G} \sum_{h=0}^{t} (r_{p,g,h} + \mu) - \sum_{g \in G} x_{p,g,t'}^s \right) \ge 0 \\
TT_{p,0}^a &= \frac{\sum_{h=0}^{T_f-1} (\upsilon_{p,0,h} - \upsilon_{p,0,h+1})(h+1)}{\sum_{g \in G} (r_{p,g,0} + \mu)} \\
TT_{p,t}^a &= \frac{\sum_{h=t}^{T_f-1} (\upsilon_{p,t,h} - \upsilon_{p,t,h+1} + \upsilon_{p,t-1,h+1} - \upsilon_{p,t-1,h})(h+1-t)}{\sum_{g \in G} (r_{p,g,t} + \mu)} \\
&\forall w \in W, g \in G, p \in P^w, t \in \mathbb{T}, t' = t, \ldots, T_f
\end{aligned} \tag{25}$$

3. Model properties

3.1. OD level first-in-first-out (FIFO) at equilibrium

As shown in Ukkusuri, Han, and Doan (2012), the cell level FIFO may not hold due to the flow updating method of diverging cells. However, the OD level FIFO is proved to hold at equilibrium for homogeneous user case (for detailed proof, see Ukkusuri, Han, and Doan 2012). We make use of this conclusion to prove the OD level FIFO for heterogeneous users.

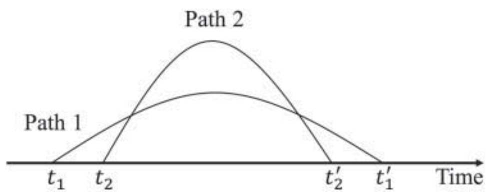


Figure 1. O–D level FIFO illustration.

In homogeneous user case, assume that there is a flow departing at time t_1 using path p_1 arriving at time t_1'; and another flow departing at time t_2 using path p_2 arriving at time t_2', where p_1 and p_2 have the same origin and destination. Ukkusuri, Han, and Doan (2012) showed that under UE condition, the following case is impossible (Figure 1):

$$t_1 < t_2, \quad t_2' < t_1'$$

Also, the proof shows that the corresponding cost for path 1, C_1, and cost for path 2, C_2, always satisfy $C_1 > C_2$. In other words, the users will always tend to choose the path that will reach the destination early, since they have smaller cost regardless of their preferred arrival time. This is a consequence of the assumption that $\beta < \alpha < \gamma$.

Now consider the heterogeneous user case, without loss of generality, consider two groups of users 1 and 2 from the same OD pair. Using the same two path examples, let f_{ij} be the flow of user group j on path i, thus $f_{i1} + f_{i2} = f_i$. Assume that we obtain a flow pattern under user equilibrium, say $f_{11} > 0, f_{21} > 0$, in which we do not have any restriction on the flow of group 2. Let C_{ij} be the cost for user group j using path i. Now consider user group 1 only, since path 1 is a longer path compared with path 2, thus we have $C_{11} > C_{21}$, however, $f_{11} \neq 0$, thus this is not possible.

Next, consider a special case that $f_{11} \neq 0, f_{21} = 0$; in other words, users of group 1 only use path 1. Since both path 1 and path 2 are valid paths for user groups 1 and 2. Thus, path 2 is also a valid path for user group 1. Now consider user group 1 only, if there is an infinitesimal amount of flow f_{21}^* on path 2, then we have $C_{11} > C_{21}'$, use the continuity of cost function C (which will be proved in Section 3.2), let $f_{21}^* \to 0$, we obtain that $C_{11} > C_{21}' = C_{21}$. Again, since $C_{11} > C_{21}$ this cannot be an equilibrium solution. Hence again, $t_1 < t_2, t_2' < t_1'$ is not possible of heterogeneous user case, and OD level FIFO holds at equilibrium conditions.

3.2. Continuity

Lemma 3.1: *Let $\mathbf{x} = CTM(\mathbf{r})$ be the function representation of cell occupancy $\mathbf{x}$ obtained from mCTM computation. The function $\mathbf{x} = CTM(\mathbf{r})$ is continuous in $\mathbf{r}$.*

Proof: As shown in Section 2.3, all CTM equations except diverging and merging flow equations are linear equations, in which continuity clearly holds. For diverging and merging cells, it suffices to show that $y_{p,g,t}^{i,j}$ is continuous in $x_{p,g,t}^{i}$. Since the continuity clearly holds when $\tilde{x}_t^{i,j} = \sum_{g\in G}\sum_{\forall p\ni(i,j)} x_{p,g,t}^{i} > 0$, we need to show that for diverging and merging cells, $\lim_{\tilde{x}_t^{i,j}\to 0} y_{p,g,t}^{i,j} = 0$.

For diverging link, from Equations (11) to (13), we have

$$0 \le y_{p,g,t}^{i,j} = \frac{x_{p,g,t}^{i}}{\bar{x}_t^i + \xi} \cdot \min\left(\tilde{x}_t^{i,j}, Q^i, Q^j, \delta(N^j - \bar{x}_t^i)\right) \cdot \min\left(1, \frac{Q^i}{\sum_{j'\in\Gamma_i}(\min(\tilde{x}_t^{i,j'}, Q^i, Q^{j'}, \delta(N^{j'} - \bar{x}_t^i))) + \xi}\right)$$

$$\le \frac{x_{p,g,t}^{i} \cdot \tilde{x}_t^{i,j}}{\sum_{g\in G}\sum_{\forall p\ni(i,j)} x_{p,g,t}^{i} + \xi}$$

And as $\tilde{x}_t^{i,j} \to 0, x_{p,g,t}^i \to 0$, thus $\lim_{\tilde{x}_t^{i,j}\to 0} y_{p,g,t}^{i,j} = 0$ holds. Similarly, for merging cell,

$$0 \le y_{p,g,t}^{i,j} = \frac{x_{p,g,t}^k}{\bar{x}_t^k + \xi} \min(Q^k, \bar{x}_t^k) \times \min\left(1, \frac{\min(Q^i, \delta(N^i - \bar{x}_t^i))}{\sum_{k' \in \Gamma_i}(\min(Q^{k'}, \bar{x}_t^{k'})) + \xi}\right) \le \frac{x_{p,g,t}^k}{\bar{x}_t^k + \xi} \min(Q^k, \bar{x}_t^k)$$

thus clearly, $\lim_{\tilde{x}_t^{i,j}\to 0} y_{p,g,t}^{k,i} \to 0$ holds.

Finally, since the cell occupancy $\boldsymbol{x}$ is continuous in $\boldsymbol{r}$, thus $\boldsymbol{x} = CTM(\boldsymbol{r})$ is continuous in $\boldsymbol{r}$. ■

Based on the definition in Section 2.4, we can show that the average path travel time $TT_{p,t}^a$ is continuous in $\boldsymbol{r}$. Denote

$$TT^a \triangleq (TT_{p,t}^a)_{p\in P, t\in \mathbb{T}}, \quad C \triangleq (C_{p,g,t})_{p\in P, g\in G, t\in \mathbb{T}}$$

Proposition 3.2: *TT^a and C are continuous in $\boldsymbol{r}$.*

Proof: We prove that this proposition follows the logic: $\boldsymbol{r} \to \boldsymbol{x} \to TT^a \to C$. As,

$$TT_{p,t}^a = \frac{\sum_{h=t}^{T_f-1}(\upsilon_{p,t,h} - \upsilon_{p,t,h+1} + \upsilon_{p,t-1,h+1} - \upsilon_{p,t-1,h})(h+1-t)}{\sum_{g\in G}(r_{p,g,t} + \mu)}, \quad \forall p \in P, t \in \mathbb{T}$$

$$\upsilon_{p,t,t'} = \max\left(0, \sum_{g\in G}\sum_{h=0}^{t}(r_{p,g,h} + \mu) - \sum_{g\in G} x_{p,g,t'}^s\right), \quad \forall t' = t, \ldots, T_f$$

By the continuity of the max function and cell occupancy $\boldsymbol{x}$, it is clear that $\upsilon_{p,t,t'}$ and $TT_{p,t}^a$ is continuous if $\sum_{g\in G}(r_{p,g,t} + \mu) > 0$ and $\sum_{g\in G}(r_{p,g,t} + \mu) \not\to 0$, more precisely, we have $TT_{p,t}^a > 0$. As $\boldsymbol{r} \ge 0$ and μ is a small positive value, hence $\sum_{g\in G}(r_{p,g,t} + \mu) > 0$ and $\sum_{g\in G}(r_{p,g,t} + \mu) \not\to 0$ always hold. Hence, $TT_{p,t}^a(\boldsymbol{r})$ is continuous in $\boldsymbol{r}$.

The total cost $C_{p,g,t}$ is given as

$$\begin{aligned} C_{p,g,t} &= \alpha_g^w TT_{p,t}^a + \beta_g^w \max(0, e_{p,g,t}) + \gamma_g^w \max(0, l_{p,g,t}) \\ &= \alpha_g^w TT_{p,t}^a + \beta_g^w \max(0, t_g^* - TT_{p,t}^a) + \gamma_g^w \max(0, TT_{p,t}^a - t_g^*) \end{aligned}$$

The continuity of $C_{p,g,t}$ is satisfied since the max function is continuous and $TT_{p,t}^a$ is continuous in $\boldsymbol{r}$. ■

Note in the above proposition, the background flow μ plays an important role in ensuring the continuity of function $TT_{p,t}^a(\boldsymbol{r})$. We show this using the following simple example illustrated in Figure 2.

Consider the simple network of one user group and two cells in Figure 2. The flow capacity out of source cell is 40, and the jam density is 500. Assuming $\mu \to 0$ and let the flow only depart at the first time step, i.e. $r_1 > 0, r_2 = 0$. We are interested in the value of average path travel time of flow that departs at the second time step TT_2^a when r_1 increases. Based on the definition of average path travel time, the plots of the average path travel time TT_1^a, TT_2^a as a function of r_1 are shown in Figure 3(a). It can be easily observed that TT_2^a is not continuous when r_1 reaches the value of an integer times the flow capacity of the cell (Q). This is because when such cases happen, the infinitesimal amount of flow

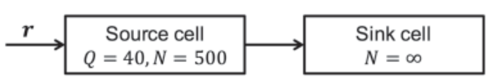


Figure 2. A simple illustration network.

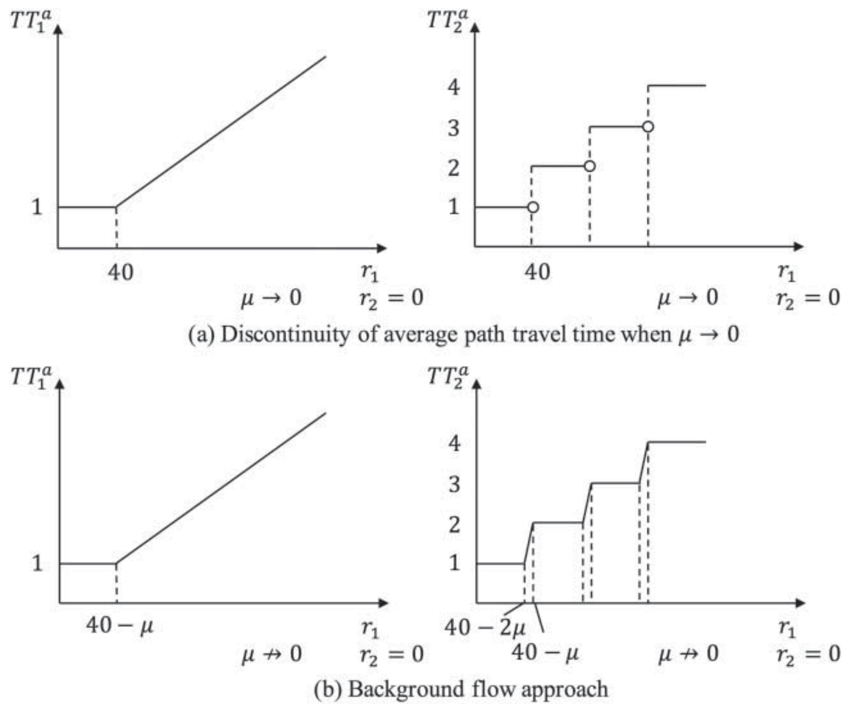


Figure 3. Illustration of the discontinuity of average path travel time when $\mu \to 0$.

μ departed at second time step will be blocked by the flow departed at the first time step, and the flow is held up for one additional time step. As $\mu \to 0, r_2 = 0$, thus $\mu + r_2 \to 0$ and

$$TT_2^a = \frac{\lim_{\mu \to 0} \mu \cdot ((t^{b,1} + 1) - 1 + 1)}{\lim_{\mu \to 0} \mu} = t^{b,1} + 1$$

where $t^{b,1}$ is the time when such blocking by r_1 occurs (when r_1/Q or r_1/N is integer). Note that the above case only occurs when $\sum_{g \in G} (r_{p,g,t} + \mu) \to 0$, which causes TT_2^a to jump from $t^{b,1}$ to $t^{b,1} + 1$ due to the blocking.

On the other hand, when μ serves a fixed positive background flow of the network, and $\mu \nrightarrow 0$. Then following Equations (16)–(18), the average path travel time of TT_1^a, TT_2^a as a function of r_1 is plotted in Figure 3(b). As the background flow is independent from the modeled flow (r_1 and r_2), under this case, the background flow μ serves as a soft buffer when such blocking occurs. As the average computation is performed on proportions of μ that arrived at $t^{b,1}$ and $t^{b,1} + 1$, the resulting average path travel time TT_2^a is still continuous.

The above example shows that the background flow μ is vital to preserve the continuity of the average path travel time TT^a and hence the path cost C. It also suggests that the path cost function is not differentiable nor smooth. The value of μ controls the level of smoothness of the cost function. As smaller μ will lead to more "spiky" cost function, that makes the problem more difficult to solve.

3.3. Non-monotonicity of the cost function

It can be easily observed that the problem is non-convex. More importantly, we can show that the monotonicity of the cost function does not hold.

Definition 3.3: A mapping $F : K \subseteq \mathbb{R}^n \to \mathbb{R}^n$ is said to be

(a) Pseudo monotone on K if for all vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in K,

$$(\boldsymbol{x}-\boldsymbol{y})^{\mathrm{T}} F(\boldsymbol{y}) \geq 0 \Rightarrow (\boldsymbol{x}-\boldsymbol{y})^{\mathrm{T}} F(\boldsymbol{x}) \geq 0 \tag{26}$$

(b) Monotone on K if

$$[F(\boldsymbol{x}) - F(\boldsymbol{y})]^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{y}) \geq 0 \tag{27}$$

The property of monotonicity is important because it closely related to the complexity involved when solving the corresponding CP/VI problem. We show the violation of this property using the same two cell networks in Figure 2. Assuming that there is a total of 200 unit flow loaded on this network with the maximum departure time to be 8. Consider following two scenarios with a single user group of user profile $\alpha = 1, \beta = 0.8, \gamma = 1.2$ and the preferred arrival time $t^* = 6$. If $\mu \ll 10$, then the associated cost can be computed as in Table 1.

From the above two scenarios, we see that

$$(\boldsymbol{r}_{\mathrm{I}} - \boldsymbol{r}_{\mathrm{II}})^{\mathrm{T}}(C(\boldsymbol{r}_{\mathrm{I}}) - C(\boldsymbol{r}_{\mathrm{II}})) = -21.79 < 0$$

indicating that the monotonicity does not hold. Moreover, even pseudo-monotonicity fails:

$$(\boldsymbol{r}_{\mathrm{I}} - \boldsymbol{r}_{\mathrm{II}})^{\mathrm{T}} C(\boldsymbol{r}_{\mathrm{II}}) = 5.79 > 0, \text{but} (\boldsymbol{r}_{\mathrm{I}} - \boldsymbol{r}_{\mathrm{II}})^{\mathrm{T}} C(\boldsymbol{r}_{\mathrm{I}}) = -16 \not> 0$$

This example shows some interesting aspects of the model. For example, in Scenario I, the trip cost in time steps 2 and 3 are higher than the cost in time step 1, even when no departure flow appears at these time steps. Also in Scenario II, shifting 10 units of flow to time step 2 decreases the cost of departing at time step 2. As the amount of flow that departs at step 2 can join the last part of remaining flow that departs at time step 1, and leave the source cell together at time step 5. We call such a phenomenon as *blocking effect*, which is caused by congestion and the use of average travel time to estimate the trip costs.

As the cell capacity Q is limited, assuming a large amount of flow f is loaded to a cell, then some proportion of it will be held in the same cell for the next time step and blocking the later arriving flows. The travel time of departing immediately after f will be higher even when departure rate is low or no departure flow at this time step, as f fully occupies the cell capacity and later flow has to wait until f is dissipated. Also, since the average travel time is used, the estimated travel time associated with f might be lower than some proportion of f actually experienced. The blocking effect means that the trip costs can be high even when the departure flow rate is low or 0 at some time step. This phenomenon causes the cost function lose many nice properties including monotonicity. Higher demand in the network will cause more severe blocking, which poses greater challenge in solving this problem.

Table 1. Two scenarios in the illustration example.

	Scenario I					Scenario II				
t	$r_t + \mu$	$\overline{TT}$	e_t	l_t	C_t	$r_t + \mu$	$\overline{TT}$	e_t	l_t	C_t
1	200	3	2	0	4.6	190	2.895	2.105	0	4.579
2	μ	5	0	1	6.2	$10+\mu$	≈ 4	0	0	≈ 4
3	μ	4	0	1	5.2	μ	4	0	1	5.2
4	μ	3	0	1	4.2	μ	3	0	1	4.2
5	μ	2	0	1	3.2	μ	2	0	1	3.2
6	μ	1	0	1	2.2	μ	1	0	1	2.2
7	μ	1	0	2	3.4	μ	1	0	2	3.4
8	μ	1	0	3	4.6	μ	1	0	3	4.6

The non-monotonicity of the cost function prohibits the use of an efficient derivative-free projection algorithm to solve this problem (Facchinei and Pang 2003b), especially for high demand case. To solve the problem, we use a modified PPA by solving a series of regularized problems that is better behaved, and eventually find the equilibrium or near equilibrium solutions. However, we first prove the solution existence for this formulation in the following section.

4. Solution existence

4.1. Preliminaries

In this section, we convert the problem into a finite-dimensional VI problem and prove the solution existence. In its general form, a VI is defined as follows.

Definition 4.1: Let set $K \in \mathbb{R}^n$ and a mapping $F : K \to \mathbb{R}^n$, the VI, denoted as VI(K, F), is to find a vector $x \in K$ such that

$$(y - x)^{\mathrm{T}} F(x) \geq 0, \quad \forall y \in K \tag{28}$$

The set of solutions to this problem is denoted SOL(K, F).

A useful result for the solution existence of VI(K, F) is given by Facchinei and Pang (2003a, Corollary 2.2.5).

Lemma 4.2: *If the set $K \subseteq \mathbb{R}^n$ is compact and convex and let $F : K \to \mathbb{R}^n$ be continuous. Then the set SOL(K, F) is nonempty and compact.*

The main idea for proving the solution existence is to show that the equivalent VI of the problem has nonempty solution set. The key step is to show the equivalency between the complementarity formulation and a VI problem. The proof uses the KKT condition for a VI, which is given as follows.

Proposition 4.3 (Facchinei and Pang 2003a, Proposition 1.3.4): *Let K be represented by finite many differentiable inequalities and equations:*

$$K \equiv \{x \in \mathbb{R}^n : h(x) = 0, g(x) \leq 0\} \tag{29}$$

with $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ being vector-valued continuously differentiable functions. Let F be a mapping from K into $\mathbb{R}^n$. The following two statements are valid:

(a) *Let $x \in$ SOL(K, F). If Abadie's CQ holds at x, then there exist vectors $\mu \in \mathbb{R}^l$ and $\lambda \in \mathbb{R}^m$ such that*

$$\begin{aligned} 0 &= F(x) + \sum_{j=1}^{l} \mu_j \nabla h_j(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) \\ 0 &= h(x) \\ 0 &\leq \lambda \perp g(x) \leq 0 \end{aligned} \tag{30}$$

(b) *Conversely, if each function h_j is affine and each function g_i is convex, and if (x, μ, λ) satisfies above conditions, then x solves the VI(K, F).*

4.2. Solution existence

Since the path cost function (denoted as $C(\boldsymbol{r})$) is continuous on $\boldsymbol{r}$, we can write the complementarity formulation in the following compact form:

$$\begin{aligned} &0 \le r_{p,g,t} \perp C_{p,g,t}(\boldsymbol{r}) - C_g^{*w} \ge 0, \quad \forall p \in P^W, g \in G, w \in W, t \in \mathbb{T} \\ &0 \le C_g^{*w} \perp \sum_{p \in P^W} \sum_{t=0}^{T} r_{p,g,t} - d_g^w \ge 0, \quad \forall w \in W, g \in G \end{aligned} \tag{31}$$

Also define

$$\Omega \triangleq \left\{ \sum_{p \in P^w} \sum_{t=1}^{T} r_{p,g,t} = d_g^w, \boldsymbol{r} \ge 0, \forall g \in G \right\} \tag{32}$$

We claim that the CP (31) is equivalent to VI(Ω, C).

Lemma 4.4: $(\boldsymbol{r}^*, C^*)$ *is a solution of Equation* (31) *if and only if it is a solution of VI*(Ω, C).

Proof: First prove sufficiency. Let $(\boldsymbol{r}^*, C^*)$ be a solution of *VI*(Ω, C), we want to show it also solves Equation (31). Since the feasible region Ω is the solution space of a system of linear equations, thus Abadie's CQ is satisfied. Let

$$h_g^w(\boldsymbol{r}) = \sum_{p \in P^w} \sum_{t=1}^{T} r_{p,g,t} - d_g^w, \quad \forall g \in G, w \in W$$

$$g_{p.g,t}(\boldsymbol{r}) = -r_{p,g,t}, \quad \forall p \in P, g \in G, t \in \mathbb{T}$$

Thus $\Omega = \{\boldsymbol{r} : h_g^w(x) = 0, g_{p,g,t}(x) \le 0, \forall w \in W, p \in P, g \in G, t \in \mathbb{T}\}$. Note that in the CP problem (31), separate $C_{p,g,t}(\boldsymbol{r}), h_g^w(\boldsymbol{r})$ and $g_{p.g,t}(\boldsymbol{r})$ are defined for each $g \in G$, by the KKT condition of a VI (Proposition 4.3), there exist $\mu^* \triangleq (\mu^{*w}_g)_{w\in W, g\in G}$ and $\lambda^* \triangleq (\lambda^*_{p,g,t})_{w \in W, g\in G}$, such that

$$C_{p,g,t}(\boldsymbol{r}^*) + \mu^{*w}_g - \lambda^*_{p,g,t} = 0, \quad \forall w \in W, p \in P, g \in G, t \in \mathbb{T}$$

$$\sum_{p \in P^w} \sum_{t=1}^{T} r_{p,g,t} - d_g^w = 0, \quad \forall g \in G, w \in W$$

$$0 \le \lambda^*_{p,g,t} \perp r_{p,g,t} \ge 0$$

Solving $\lambda^*_{p,g,t}$ from the first equation, we have

$$\begin{aligned} &0 \le r_{p,g,t} \perp C_{p,g,t}(\boldsymbol{r}^*) + \mu^{*w}_g \ge 0 \\ &\sum_{p \in P^w} \sum_{t=1}^{T} r_{p,g,t} - d_g^w = 0, \quad \forall g \in G, w \in W \end{aligned} \tag{33}$$

Since $C_{p,g,t}(\boldsymbol{r}^*) \ge 0$, if $\mu^{*w}_g > 0$, then $C_{p,g,t}(\boldsymbol{r}^*) + \mu^{*w}_g > 0$. From the complementarity condition in Equation (33), we have $r_{p,g,t} = 0$, hence

$$\sum_{p \in P^w} \sum_{t=1}^{T} r_{p,g,t} = 0 < d_g^w, \quad \forall g \in G, w \in W$$

which is a contradiction. Thus $\mu^{*w}_g \le 0$, let $C_g^{*w} = -\mu^{*w}_g$, we see that $(\boldsymbol{r}^*, C^*)$ solves Equation (31).

Next, we show the necessity. Let $(\boldsymbol{r}^*, C^*)$ be a solution of Equation (31), we first show

$$\sum_{p\in P^w}\sum_{t=1}^{T} r_{p,g,t} - d_g^w = 0, \quad \forall g \in G, w \in W$$

Since $\sum_{p\in P^w}\sum_{t=0}^{T} r_{p,g,t} - d_g^w \geq 0$, suppose $\sum_{p\in P^w}\sum_{t=1}^{T} r_{p,g,t} > d_g^w$ for some $w \in W, g \in G$, then from the second complementarity condition in Equation (31), we have $C_g^{*w} = 0$. As $r_{p,g,t} > 0$ for some $p \in P^w, g \in G$, which requires $C_{p,g,t}(\boldsymbol{r}) - C_g^{*w} = 0$, thus $C_{p,g,t}(\boldsymbol{r}) = 0$. This is a contradiction, since $r_{p,g,t} > 0$ requires $C_{p,g,t}(\boldsymbol{r}) > 0$. Consequently, following equation holds:

$$\sum_{p\in P^w}\sum_{t=1}^{T} r_{p,g,t} - d_g^w = 0, \quad \forall g \in G, w \in W$$

Now, we show $(\boldsymbol{r}^*, C^*)$ is indeed a solution of VI(Ω, C). We have already shown that $\boldsymbol{r}^* \in \Omega$, and again let

$$h_g^w(\boldsymbol{r}) = \sum_{p\in P^w}\sum_{t=1}^{T} r_{p,g,t} - d_g^w, \quad \forall g \in G, w \in W$$

$$g_{p.g,t}(\boldsymbol{r}) = -r_{p,g,t}, \quad \forall p \in P, g \in G, t \in \mathbb{T}$$

Notice that $h_g^w(\boldsymbol{r})$ is affine and $g_{p,g,t}(\boldsymbol{r})$ is convex, and $h_g^w(\boldsymbol{r}) = 0$ is already proved and $g_{p.g,t}(\boldsymbol{r}) \leq 0$ is given, thus by Proposition 4.3(b), $(\boldsymbol{r}^*, C^*)$ is a solution of VI(Ω, C) if the KKT condition of VI(Ω, C), Equation (33) is satisfied.

As we have already showed that the second equation holds, we only need to show the first complementarity condition holds. Note that by replacing C^{*w} with $-\mu_g^{*w}$, in the first complementarity condition of Equation (31), we obtain exactly the same complementarity condition in Equation (33). Thus, the solution of Equation (31) is a solution of VI(Ω, C), therefore the necessity condition holds. ■

Theorem 4.5: *The CP* (31) *has a solution.*

Proof: From Lemma 4.2, the problem VI(Ω, C) has nonempty solution set if Ω is a compact convex set. Since

$$\Omega \triangleq \left\{\sum_{p\in P^w}\sum_{t=1}^{T} r_{p,g,t} = d_g^w, \boldsymbol{r} \geq 0, \forall g \in G\right\}$$

which is a system of linear equalities, thus clearly, Ω is closed and convex. Also since

$$0 \leq r_{p,g,t} \leq \sum_{p\in P^w}\sum_{t=1}^{T} r_{p,g,t} = d_g^w$$

Thus, Ω is bounded. Since Ω is compact and convex, SOL(Ω, C) is nonempty. As we have proved in Lemma 3.6, the solution of VI(Ω, C) also solves Equation (31). Consequently, the CP (31) has nonempty solution set. ■

5. Solution algorithm

Due to the complexity and non-monotonicity of the problem, commercial solvers such as PATH or KNITRO fail to solve the problem. Thus, a more robust solution approach is needed. Ukkusuri, Han, and Doan (2012) used the projection algorithm with variable steps (PAVS) (Facchinei and Pang 2003b, Algorithm 12.1.4) to obtain the solution of the equivalent VI for a single user class DUE model with embedded path-based CTM. An obvious weakness of the projection algorithm, as pointed out in Facchinei and Pang (2003b), lies in the requirement of strongly monotonicity. This does not hold even for the single user class model as discussed in Section 3.3. Consequently, the projection algorithm is not guaranteed to converge to the equilibrium solution. To obtain a convergent solution, Ukkusuri, Han, and Doan (2012) suggested to shrink the magnitude of the updating scale parameter in PAVS to force the algorithm return a convergent solution. But there is no concrete criterion given on how to properly shrink the scale parameter, nor proof provided that the forced convergent algorithm indeed solves the VI problem.

The major difficulty of solving the equivalent VI problem VI(Ω, C) lies in the non-differentiability and non-monotonicity of the cost function C. In this study, a modified PPA is used to find the equilibrium or close to equilibrium solution of the problem. The PPA is developed by Rockafellar (1976) and is one of the most popular methods for solving nonlinear equations, optimization problems as well as the VI problems. The key idea of PPA is to iteratively solving a series of perturbed sub-problems which adds a strongly monotone regularization term $\tau(I - x^n)$. Hence if the original problem is monotone, then the resulting sub-problems will be strongly monotone and can be easily solved using the conventional projection algorithm. Although the global convergence of PPA requires the monotonicity of the original problem (Rockafellar 1976; Facchinei and Pang 2003b), several researchers have suggested that the monotonicity might not be needed for local convergence (Eckstein and Ferris 1999; Pennanen 2002) and successfully applied PPA on solving non-monotone problems with special structures. Based on these nice features of PPA, we develop a modified version of PPA to solve the multiclass traffic assignment problem. The detailed algorithm is presented as follows:

Algorithm 5.1: Modified PPA

Step 0: Set $n = 0$. Initialize a feasible departure rate $\boldsymbol{r}^0$, let $\{\rho_n\}_{n=0}^{\infty}$ and $\tau > 0$ be given.
Step 2: Solve the regularized sub-problem: $VI(\Omega, C + \tau(I - x^n))$ using projection algorithm:
  Let $\boldsymbol{z}^0 = \boldsymbol{r}^n$.
  For $k = 1$ to maximum projection iteration:
    Step 2.1: Run simulation $CTM(\boldsymbol{z}^{k-1})$, compute cost $C(\boldsymbol{z}^{k-1})$
    Step 2.2: Solve

$$\boldsymbol{z}^k = \Pi_\Omega[\boldsymbol{z}^{k-1} - \lambda(C(\boldsymbol{z}^{k-1}) + \tau(\boldsymbol{z}^{k-1} - \boldsymbol{r}^n)]$$

    Step 2.3 (early exit): For $k > 1$,
    If $\|\boldsymbol{z}^k - \boldsymbol{z}^{k-1}\| > 1.1 \cdot \|\boldsymbol{z}^{k-1} - \boldsymbol{z}^{k-2}\|$, exit the loop and return $\boldsymbol{z}^* = \boldsymbol{z}^{k-1}$. Otherwise, if $k =$ maximum projection iteration, return $\boldsymbol{z}^* = \boldsymbol{z}^k$
Step 3: Set

$$\boldsymbol{r}^{n+1} \triangleq (1 - \rho_n)\boldsymbol{r}^n + \rho_n \boldsymbol{z}^*$$

Step 4: Convergence check:
  If $\boldsymbol{r}^{n+1} - \boldsymbol{r}^n < \epsilon$, terminate the algorithm, $\boldsymbol{r}^* = \boldsymbol{r}^{n+1}$. Otherwise $n = n + 1$, go to step 2.

---

where $\Pi_\Omega(\boldsymbol{x})$ denotes the projection of $\boldsymbol{x}$ onto set Ω, mathematically solved as a quadratic programming problem: $\Pi_\Omega(\boldsymbol{x}) = \text{argmin}_{\boldsymbol{y}\in\Omega}(1/2)(\boldsymbol{y} - \boldsymbol{x})^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{x})$.

The modified PPA uses the same principle as typical PPA, that iteratively solves regularized sub-problems VI$(\Omega, C + \tau(I - x^n))$. However, to guarantee convergence, PPA usually requires the solution from the sub-problem has to be exact, or inexact is allowed but the error is bounded within some tolerance criteria (Rockafellar 1976), which unfortunately, cannot be satisfied for this problem. Although by choosing a sufficiently large τ value, it is possible that the sub-problems VI$(\Omega, C + \tau(I - x^n))$ will become monotone, it is not always guaranteed. Consequently, the projection algorithm that used to solve the sub-problem may fail to converge. To alleviate such difficulties, we suggest an additional early exit condition (Step 2.3 in Algorithm 5.1), which stops the projection algorithm in the middle when we observe a possible divergence occur $((\|\boldsymbol{z}^k - \boldsymbol{z}^{k-1}\| > 1.1 \cdot \|\boldsymbol{z}^{k-1} - \boldsymbol{z}^{k-2}\|)$. This will allow us to obtain a relative close solution of $\mathrm{VI}(\Omega, C + \tau(I - x^n))$ and hence lead to better performance of the algorithm. The numerical experiments also show that this strategy can greatly speed up the convergence and produces superior solutions.

In the next section, we present the numerical results obtained using the modified PPA.

6. Numerical experiments

The DUE problem as well as the solution approach are coded and tested in MATLAB. The multiclass path-based CTM module are compiled into C codes to improve the computation efficiency. Extensive experiments are conducted on three test networks (namely, X-shape, Nguyen–Dupuis and Sioux Falls network) under different traffic conditions. All the numerical experiments are performed on an Intel i7 CPU desktop. For detailed set-up of algorithm parameters, we use $\lambda = 0.25$, $\rho_n = 1.7 \times (n+1)^{0.8}$ and the maximum projection iteration is set to 50 (40 for Sioux Falls network). For the convergence criteria, we use $\epsilon = 0.05$ for all tests. For most of the experiments, we consider two user groups: one group with lower arrival and late penalty ($\alpha = 1, \beta = 0.5, \gamma = 2$) and relatively late preferred arrival time, which correspond to users with higher travel flexibility; the other group corresponds to users with higher penalty ($\alpha = 1, \beta = 0.6, \gamma = 2.5$) and relatively early preferred arrival time. In addition, a test with three user groups is also conducted on X-shape network. Different demand scenarios are tested. Specifically, we consider both low demand level (low congestion level, total demand $= d$) and high demand level (heavily congested, total demand $= 2d$ or *2.5d*). As discussed in previous sections, the conventional projection algorithm fails to solve the equivalent VI problem, while the modified PPA approach produces reasonably good solutions.

To further investigate the quality of the solution obtained, we introduce following penalty function to evaluate the relative gap between the current solution and the equilibrium solution:

$$P(\boldsymbol{r}, C(\boldsymbol{r})) = \sum_{i \in I} [C_i(\boldsymbol{r}) - \min\,(C_i(\boldsymbol{r}))]^{\mathrm{T}} r_i \tag{34}$$

in which we partition $\boldsymbol{r}$ and $C(\boldsymbol{r})$ into a set of groups indexed by a particular user group and OD pair i. The penalty function is always non-negative and $P(\boldsymbol{r}^*, C(\boldsymbol{r}^*)) = 0$ only when $\boldsymbol{r}^* \in \mathrm{SOL}(C, \Omega)$. The penalty measures the degree of deviation to the equilibrium solution. As at equilibrium, all flow of a particular user group and OD pair departs only when the path and time step yields minimum cost $(\min\,(C_i(\boldsymbol{r})))$. Hence for nonequilibrium solution, the greater amount of flow that deviates from the minimum cost paths and departure times, the larger penalty value will be. As the global convergence criteria of the PPA for the optimal solution cannot be established, we use the penalty function to evaluate the quality of the obtained solution.

In all experiments, the proposed solution approach successfully finds close to equilibrium solutions. Simultaneous route and departure time choice behaviors are observed in the results. Solutions are summarized in Table 2 and presented in Figures 5–7. The performance of the algorithm is further investigated in Section 6.3.

6.1. Test networks and scenarios

The cell representation of the three networks is presented in Figure 4. The summary for the test networks and scenarios are as follows:

(1) An X-shape network with 10 cells: two origins and two destinations cells. No route choice in this network. Both two user groups and three user groups are tested. The demand scenarios include a low demand case (total demand $d = 320$) and a high demand case ($2d = 640$).

Table 2. Results of the numerical experiments.

#	Network	Groups	Total Demand	Iterations	Running time (s)	Penalty
1	X-shape	2	320	176	117.6	49.99
2	X-shape	2	640	42	21.5	120.89
3	X-shape	3	320	311	354.9	11.78
4	Nguyen–Dupuis	2	440	200	1549.3	107.14
5	Nguyen–Dupuis	2	1100	132	625.5	580.18
6	Sioux Falls	2	585	91	1047.4	60.80
7	Sioux Falls	2	1170	66	659.35	270.93

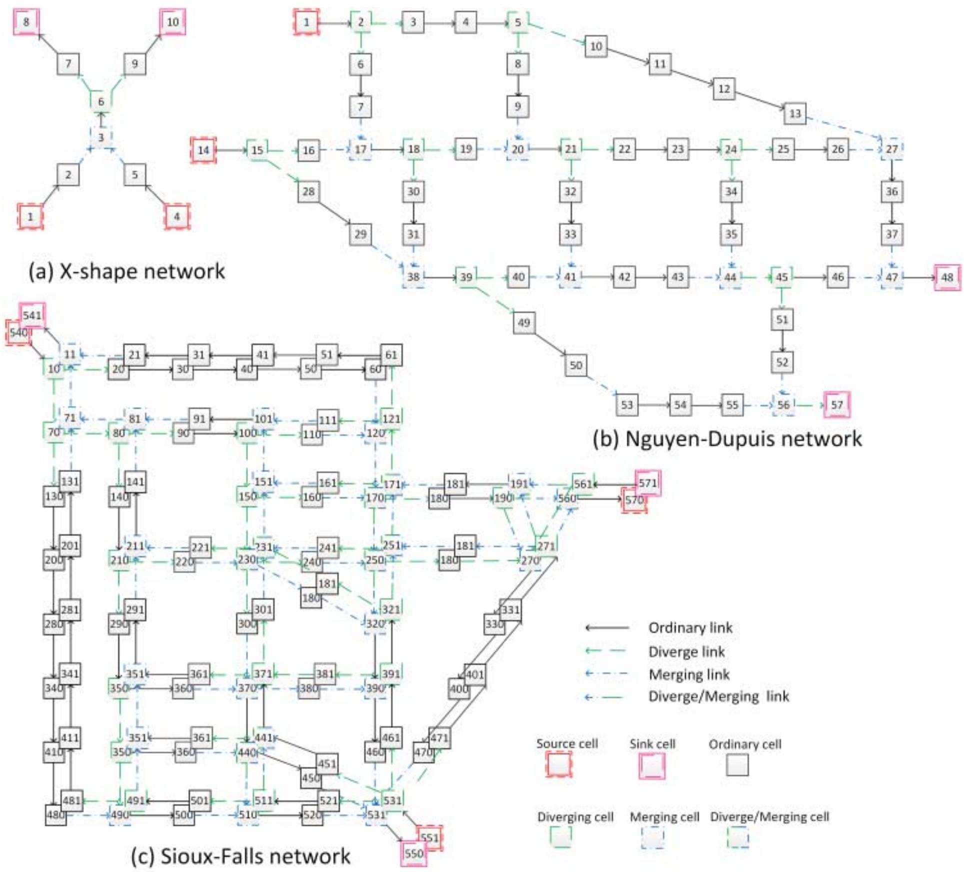


Figure 4. Test networks.

(2) The network of Nguyen and Dupuis (1984) in cell representation, containing 57 cells and 63 links. There are two origins (cells 1 and 14) and two destinations (cell 48 and 57), constituting 4 OD pairs and 12 paths. The demand scenarios include a low demand case (total demand $d = 440$) and a high demand case ($2.5d = 1100$).

(3) The Sioux Falls network with 114 cells and 142 links. There are six OD pairs 540–550, 551–541, 540–570, 551–570, 571–541, 571–550. A number of paths are predefined for each OD pair, with two OD pairs (540–550, 551–541) with three paths, two OD pairs with two paths (540–570, 571–541) and two OD pairs (551–570, 571–550) with one path. The low demand scenario has a total demand of $d = 585$ and a high demand scenario has a total demand of $2d = 1170$.

6.2. Numerical results

6.2.1. X-shape network

The solutions for two and three user groups with low demand cases and two user groups with high demand case on X-shape network are shown in Figure 5. High-quality solutions are obtained for all tests with very low penalty values (Table 2). The three user groups with low demand case even produce a solution with a penalty value of 11.78. Considering the total amount of flow (320) and the minimum cost (about 7), as well as the possible numerical inaccuracy (largely due to the introduction of background flow) while running the CTM simulation, the penalty value is very low (at most 1.68 unit of flow violates the equilibrium condition). The results in Figure 5 show that for all tests, almost exact

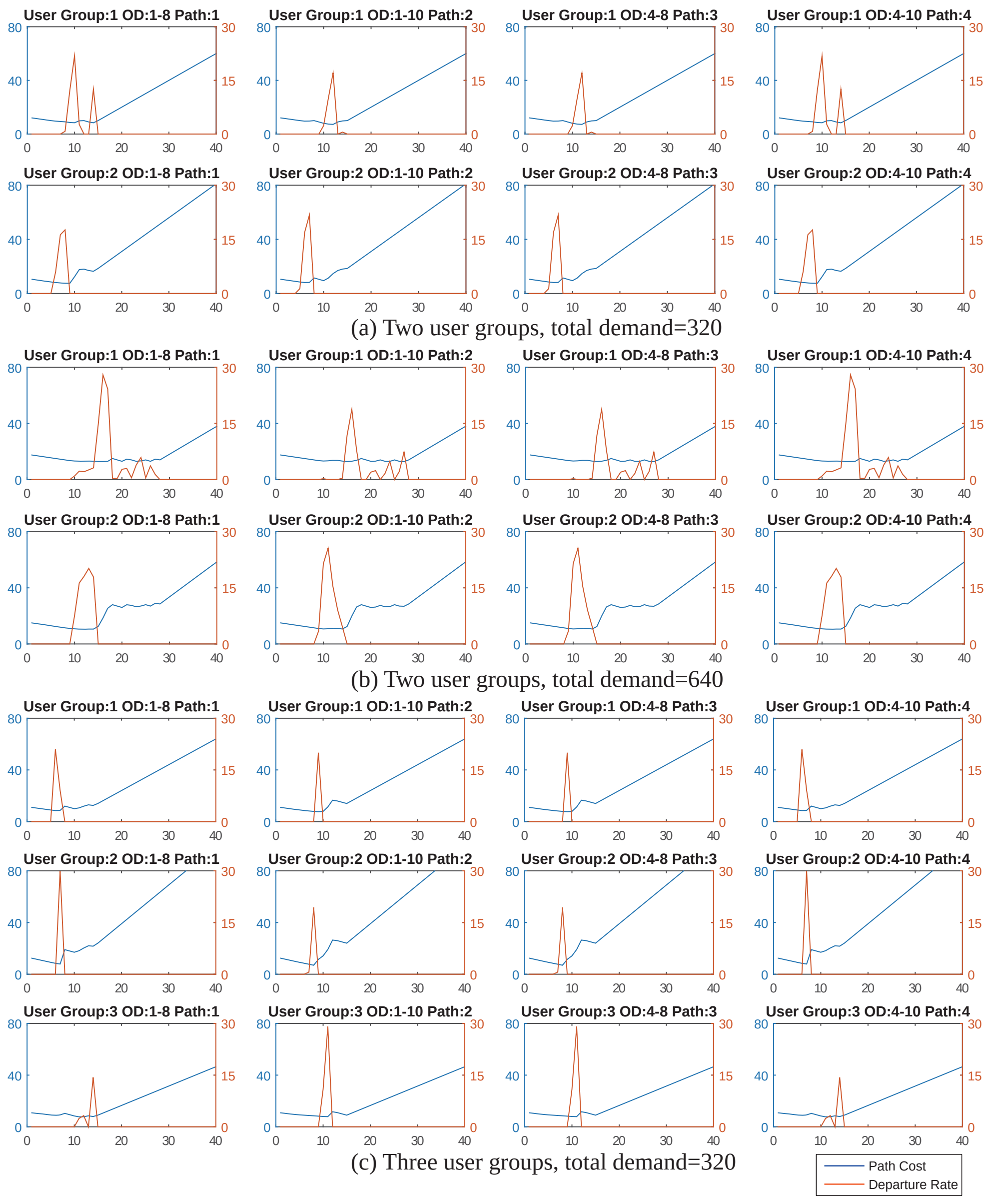


Figure 5. Solutions of X-shape network.

equilibrium solutions are obtained in low demand cases, as all flows are departing at lowest cost time step. In high demand case, majority of the departure flows are found to depart at minimum cost time step. Slight violation of equilibrium condition is observed, but the general flow pattern is very close to equilibrium condition. The second user group in both the two- and three-group tests are the users having higher late arrival penalty and earlier preferred arrival time. This user preference is reflected in the results, that in all tests, the user group 2 departs earlier than other groups.

6.2.2. Nguyen–Dupuis network

The solutions of Nguyen–Dupuis network tests are presented in Figure 6. Again, majority of the flows are departing at lowest cost time steps, and the costs of used paths for the same user group and OD pair are the same. Similar to the results in X-shape network, almost exact equilibrium solutions are obtained for low demand case and heterogeneous user travel behaviors are observed. User group 2 with higher early and late arrival penalty and earlier preferred arrival time is found to depart early for both high and low demand cases. Furthermore, although 12 paths are available in Nguyen–Dupuis network, only 4 paths are utilized for low demand case. If compare the solutions for low and high demand cases, it

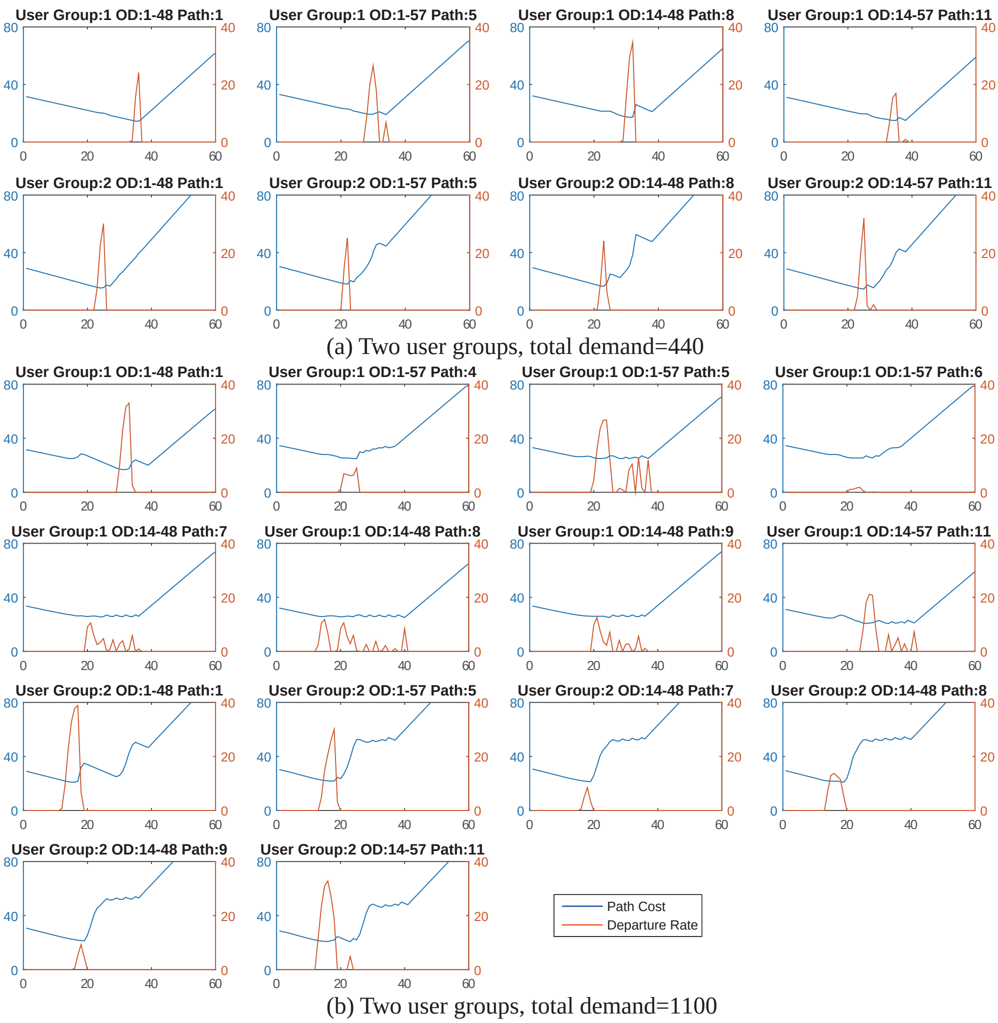


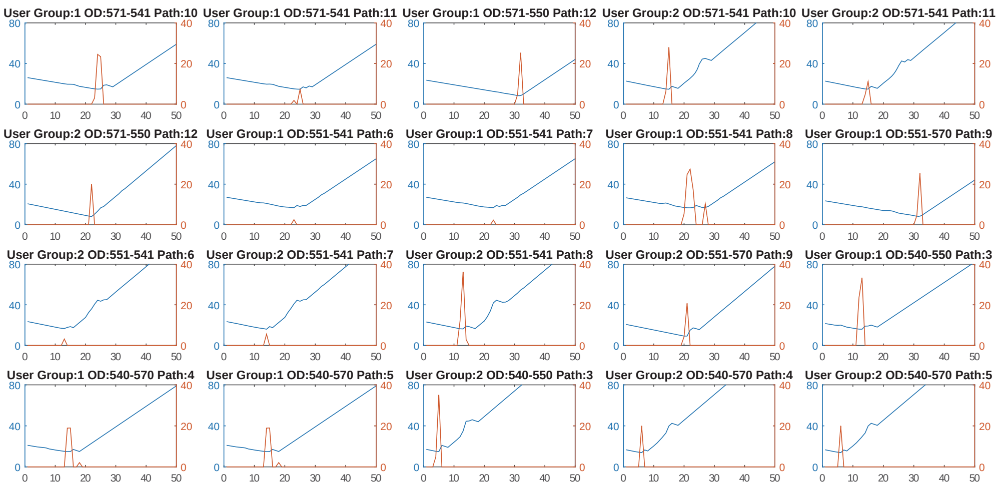
Figure 6. Solutions of Nguyen–Dupuis network.

can be observed that with the increase in total demand, more paths are utilized to reduce the path costs. For example, paths 4, 6, 7, 9 for user group 1 and paths 7, 9 for user group 2 become feasible paths although they are not used in low demand case. User group 1 utilized more paths because it has higher total demand (600 units flow) than user group 2 (500 units flow). This suggests that the route choice behavior can be effectively captured in the model.

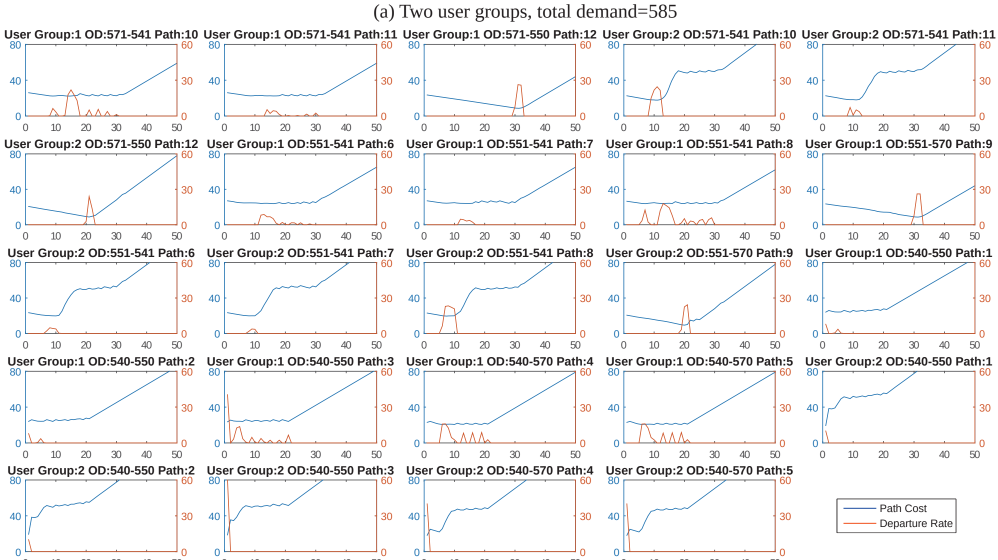
6.2.3. Sioux falls network

The solutions for Sioux Falls network tests are presented in Figure 8. As the Sioux Falls network is much larger than previous networks and the CTM simulation is costly, we set the maximum projection iteration to be 40. This will potentially decrease the solution precision of the sub-problems of PPA, but will improve the computation efficiency. Both low and high demand cases yield small penalty values (60.8 and 270.93), which suggest close to equilibrium solutions are obtained.

For the results in Figure 7, most of flows are departing at lowest cost time step, and the solutions match well with the equilibrium condition especially for the low demand case. It is observed that the



(a) Two user groups, total demand=585



(b) Two user groups, total demand=1170

Figure 7. Solutions of Sioux Falls network.

departure rate and the cost patterns are more complex due to more opportunities of flow interaction in larger network. User group 1 is designed as users with lower early and late arrival penalty and more flexible arrival time, who also have a larger share in total demand (330 and 660 units of flow for low and high demand case) compared with user group 2 (235 and 470 units of flow for low and high demand case). In the results of both low and high demand tests, travelers of user group 2 tend to depart early. Furthermore, a comparison between Figure 7(a) and (b) shows that with the increase in demand, travelers of user group 2 will depart even earlier to avoid more severe congestion and arrive on time. This is a very realistic reflection of travelers' commuting behavior during peak hours. Route choice behaviors are also captured, travelers from the same user group and OD pair only select paths and corresponding departure time that have lowest or equal cost. Similar to results in the Nguyen–Dupuis network, with the increase in demand, more previously high cost paths become feasible paths and utilized by both user groups.

6.3. Convergence and running time

All test results for different networks and demand scenarios are summarized in Table 2. One potential drawback of PPA is that it is relatively slow to solve large problems as we are solving a series of perturbed sub-problems at each iteration. And each sub-problem itself will take a relatively long time to solve using the projection algorithm when the network is large. In Table 2, solving the Nguyen–Dupuis network with low demand takes about 25 min using 200 iterations. The computation time can potentially become longer for a larger and more complex network.

It is observed that the results of low demand scenarios typically have higher quality with lower penalty values, while slight violations of DUE condition are observed for some high demand scenario tests. This is because that high demand scenarios have more severe congestion, causing more flow blocking phenomena discussed in Section 3.3. This leads to a higher level of non-monotonicity of the cost function and makes the problem harder to solve.

The convergence plots of Nguyen–Dupuis network and Sioux Falls network are shown in Figure 8(a). It can be observed that PPA converges very fast in the first few iterations, but converges slowly in later iterations. This might due to several reasons. First, as the problem is non-monotone, it is not guaranteed to find the exact solution of the perturbed sub-VI problem of PPA, which can lead to potential inefficiencies in converging to the optimal solutions; second, PPA solves a series of perturbed problems rather than the original one, hence when the current solution is close to the optimal solution, the perturbation itself may introduce interferences and slow down the convergence speed. Figure 8(b) shows the penalty value of Nguyen–Dupuis network and Sioux Falls network tests at each iteration. It can be observed that PPA indeed improves the quality of the solution (decreasing penalty

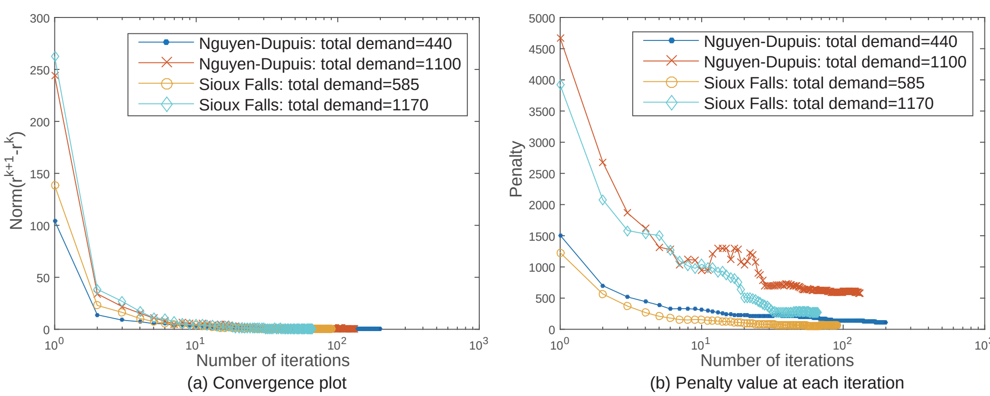


Figure 8. Convergence and penalty plots of Nguyen–Dupuis and Sioux Falls networks.

values) as the number of iteration increases even when the problem is non-monotone. In this study, we only used very loose stopping criteria ($\epsilon = 0.05$). It can be expected that when using a smaller threshold value ϵ, the quality of the solution can be further improved, however, at the cost of longer computation time.

7. Conclusions

We developed a complementarity formulation for a multi-user class, simultaneous route and departure time choice DUE model. Several model properties and the solution existence are proved. Three test networks: X-shape network, Nguyen–Dupuis network and Sioux Falls network are tested and solved using a modified PPA. The numerical results show that the solution approach can find equilibrium or close to equilibrium solutions of the problem. The main contributions of this work include:

- The development of a path-based mCTM for general networks. mCTM can serve as an embedded network loading model that capturing spatial queuing and other flow propagation properties.
- Proposing a new formulation for multi-user class DUE model, which allows for simultaneous route and departure time choice behaviors. Compared with other multiclass DTA models proposed in the literature, the new model is applicable to general network, and can be effectively solved.
- A modified PPA is proposed to solve the complex non-monotone equivalent VI problem. The results show that the proposed algorithm effectively finds equilibrium or close to equilibrium solutions.

Currently, the PPA-based solution approach still suffers from slow convergence and relatively long computation time for large networks. Future research can be done to utilize more properties of the multiclass DUE model (e.g. the Cartesian product structure of the feasible region Ω) to develop more efficient and distributed algorithms to solve the proposed DTA problem.

Acknowledgements

The authors are grateful for the support. The findings are the responsibility of the authors alone.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research is supported by National Science Foundation (NSF) [grant number 1017933].

References

Arnott, R., A. de Palma, and R. Lindsey. 1988. “Schedule Delay and Departure Time Decisions with Heterogeneous Commuters.” *Transportation Research Record* 1197: 56–57.

Arnott, R., A. de Palma, and R. Lindsey. 1994. “The Welfare Effects of Congestion Tolls with Heterogeneous Commuters.” *Journal of Transport Economics and Policy* 28 (2): 139–161.

Bliemer, M. C. J. 2007. “Dynamic Queuing and Spillback in Analytical Multiclass Dynamic Network Loading Model.” *Transportation Research Record: Journal of the Transportation Research Board* 2029: 14–21.

Bliemer, M. C. J., and P. H. L. Bovy. 2003. “Quasi-variational Inequality Formulation of the Multiclass Dynamic Traffic Assignment Problem.” *Transportation Research Part B: Methodological* 37 (6): 501–519.

Cohen, Y. 1987. “Commuter Welfare Under Peak-Period Congestion Tolls: Who Gains and who Loses?” *International Journal of Transport Economics* 14 (3): 238–266.

Dafermos, S. 1972. “The Traffic Assignment Problem for Multiclass User Transportation Networks.” *Transportation Science* 6: 73–87.

Daganzo, C. F. 1994. “The Cell Transmission Model: a Simple Dynamic Representation of Highway Traffic.” *Transportation Research Part B: Methodological* 28 (4): 269–287.

Daganzo, C. F. 1995. “The Cell Transmission Model, Part ii: Network Traffic.” *Transportation Research Part B: Methodological* 29 (2): 79–93.

Doan, K., and S. V. Ukkusuri. 2012. "On the Holding-Back Problem in the Cell Transmission Based Dynamic Traffic Assignment Models." *Transportation Research Part B: Methodological* 46 (9): 1218–1238.
Eckstein, J., and M. C. Ferris. 1999. "Smooth Methods of Multipliers for Complementarity Problems." *Mathematical Programming* 86: 65–90.
Facchinei, F., and J. Pang. 2003a. *Variational Inequalities and Complementarity Problems*. Volume I. New York: Springer-Verlag.
Facchinei, F., and J. Pang. 2003b. *Variational Inequalities and Complementarity Problems*. Volume II. New York: Springer-Verlag.
Han, L., S. Ukkusuri, and K. Doan. 2011. "Complementarity Formulations for the Cell Transmission Model Based Dynamic User Equilibrium with Departure Time Choice, Elastic Demand and User Heterogeneity." *Transportation Research Part B: Methodological* 45 (10): 1749–1767.
Liu, Y., and Y. Nie. 2011. "Morning Commute Problem Considering Route Choice, User Heterogeneity and Alternative System Optima." *Transportation Research Part B: Methodological* 45 (4): 619–642.
Lo, H. K., B. Ran, and B. Hongola. 1996. "Multiclass Dynamic Traffic Assignment Model: Formulation and Computational Experiences." *Transportation Research Record: Journal of the Transportation Research Board* 1537 (1): 74–82.
Lo, H. K., and W. Y. Szeto. 2002. "A Cell-Based Variational Inequality Formulation of the Dynamic User Optimal Assignment Problem." *Transportation Research Part B: Methodological* 36 (5): 421–443.
Merchant, D. K., and G. L. Nemhauser. 1978. "A Model and an Algorithm for the Dynamic Traffic Assignment Problems." *Transportation Science* 12 (3): 183–199.
Mesa-Arango, R., and S. V. Ukkusuri. 2014. "Modeling the car-Truck Interaction in a System-Optimal Dynamic Traffic Assignment Model." *Journal of Intelligent Transportation Systems* 18 (4): 327–338.
Nguyen, S., and C. Dupuis. 1984. "An Efficient Method for Computing Traffic Equilibria in Networks with Asymmetric Transportation Costs." *Transportation Science* 18 (2): 185–202.
Pang, J. S., L. Han, G. Ramadurai, and S. Ukkusuri. 2012. "A Continuous-Time Linear Complementarity System for Dynamic User Equilibria in Single Bottleneck Traffic Flows." *Mathematical Programming* 133 (1–2): 437–460.
Peeta, S., and A. K. Ziliaskopoulos. 2001. "Foundations of Dynamic Traffic Assignment: The Past, the Present and the Future." *Networks and Spatial Economics* 1 (3): 233–265.
Pennanen, T. 2002. "Local Convergence of the Proximal Point Algorithm and Multiplier Methods Without Monotonicity." *Mathematics of Operations Research* 27 (1): 170–191.
Ramadurai, G., S. V. Ukkusuri, J. Zhao, and J.-S. Pang. 2010. "Linear Complementarity Formulation for Single Bottleneck Model with Heterogeneous Commuters." *Transportation Research Part B: Methodological* 44 (2): 193–214.
Ran, B., D. Lee, and M. Shin. 2002. "New Algorithm for a Multiclass Dynamic Traffic Assignment Model." *Journal of Transportation Engineering* 128 (4): 323–335.
Rockafellar, R. T. 1976. "Monotone Operators and the Proximal Point Algorithm." *SIAM Journal on Control and Optimization* 14: 877–898.
Szeto, W. Y., Y. Jiang, and A. Sumalee. 2011. "A Cell-Based Model for Multi-Class Doubly Stochastic Dynamic Traffic Assignment." *Computer-Aided Civil and Infrastructure Engineering* 26 (8): 595–611.
Szeto, W. Y., and H. K. Lo. 2004. "A Cell-Based Simultaneous Route and Departure Time Choice Model with Elastic Demand." *Transportation Research Part B: Methodological* 38 (7): 593–612.
Tuerprasert, K., and C. Aswakul. 2010. "Multiclass Cell Transmission Model for Heterogeneous Mobility in General Topology of Road Network." *Journal of Intelligent Transportation Systems* 14 (2): 68–82.
Ukkusuri, S. V., L. Han, and K. Doan. 2012. "Dynamic User Equilibrium with a Path Based Cell Transmission Model for General Traffic Networks." *Transportation Research Part B: Methodological* 46 (10): 1657–1684.
Vickrey, W. S. 1969. "Congestion Theory and Transport Investment." *American Economic Review* 59: 251–261.