

Capacitated transit service network design with boundedly rational agents

Jiangtao Liu

School of Sustainable Engineering and the Built Environment,
Arizona State University, Tempe, AZ, 85281, USA

Email: jliu215@asu.edu

Xuesong Zhou

School of Sustainable Engineering and the Built Environment,
Arizona State University, Tempe, AZ, 85281, USA

Email: xzhou74@asu.edu

Tel.: +1 480 9655827

(Corresponding Author)

Submitted for publication in Transportation Research Part B

Abstract

This paper proposes a new alternative modeling framework to systemically account for boundedly rational decision rules of travelers in a dynamic transit service network with tight capacity constraints. Within a time-discretized space-time network, the time-dependent transit services are characterized by traveling arcs and waiting arcs with constant travel times. Instead of using traditional flow-based formulations, an agent-based integer linear formulation is proposed to represent boundedly rational decisions under strictly imposed capacity constraints, due to vehicle carrying capacity and station storage capacity. Focusing on a viable and limited sets of space-time path alternatives, the proposed single-level optimization model can be effectively decomposed to a time-dependent routing sub-problem for individual agents and a knapsack sub-problem for service arc selections through the Lagrangian decomposition. In addition, several practically important modeling issues are discussed, such as dynamic and personalized transit pricing, passenger inflow control as part of network restraint strategies, and penalty for early/late arrival. Finally, numerical experiments are performed to demonstrate the methodology and computational efficiency of our proposed model and algorithm.

Keywords: Dynamic transit service network design; Boundedly rational agents; Tight capacity constraint; Agent-based model

1. Introduction

In general, the transportation network design problem aims to minimize the total transportation system disutility by optimizing the location of capacity enhancement strategies or various transportation service plans. Specifically, the aim of the transit service network design problem is to provide better service to users and to increase operating efficiency, involving key decisions such as constructing a new transit line, adding train or bus schedules, or dynamically selecting different types of vehicles to meet time-dependent transit demand.

While a large number of existing studies have been devoted to network design problems with static origin-destination (OD) demand input, this paper intends to study a class of practically important problems for designing discrete transit service networks with (i) individual traveler agents (corresponding to time-dependent demand matrices) and (ii) tight time-varying capacity constraints in terms of the number of passengers a transit vehicle or a station can carry. In addition, our study aims to address a number of theoretically challenging questions for realistically capturing and possibly affecting individual traveler's behavior in an oversaturated transit system. Under possibly extremely heavy congested conditions in transit systems (e.g. in Beijing and Tokyo), for example, each traveler wants to minimize his/her disutility within his/her rational/tolerance bands and preferred arrival times, but finally they need to select a close-to-user-optimal path from a limited number of capacity-feasible routing options, which could have very different path travel times and path-dependent prices.

1.1 Transportation network design problem

The discrete transportation network design problem has been traditionally formulated as a bi-level programming model, where the upper-level problem decides where and how many links should be built, and the lower-level problem aims to predict travelers' response to changes in the network conditions, by assuming certain user behavior rules such as Wardrop or Nash user equilibrium. A number of comprehensive reviews on the discrete traffic network design problem have been offered by the classical papers by [Magnanti and Wong \(1984\)](#), [Yang and Bell \(1998\)](#), as well as a recent study by [Farahani et al. \(2013\)](#). Typical solution algorithms in the discrete cases include branch and bound ([Leblanc, 1974](#)), support functions in bender decomposition ([Gao et al., 2005](#)), meta-heuristics (e.g. [Xiong and Schneider, 1992](#); [Drezner and Wesolowsky, 2003](#); [Poorzahedy and Rouhani, 2007](#)).

Another active research line for considering user equilibrium (UE) conditions is to construct a single-level programming model with constraints corresponding to the UE principle. The early work by [Bard and Moore \(1990\)](#) provided a reformulation based on the Karush–Kuhn–Tucker (KKT) conditions with a branch and bound solution scheme. Recently, [Farvaresh and Sepehri \(2011\)](#) presented a single-level mixed integer linear problem (MILP) by representing the UE condition as KKT constraints and linearized those non-linear terms by introducing binary auxiliary variables. In addition, [Lou et al. \(2009\)](#) studied a robust network design approach and formulated the problem as a mathematical program with complementarity constraints, where a cutting-plane scheme was proposed for solving this problem with demand uncertainties. [Wang and Lo \(2010\)](#) recast the lower-level model by complementarity constraints, and then introduced a set of binary variables to transform the “if-then” conditions into an equivalent set of linear constraints, which can be further approximated with a MILP framework to search for global optimal solutions. [Luathep et al. \(2011\)](#) formulated the user equilibrium condition as a variational inequality problem, and also adopted the linear approximation technique to transform the original problem as an MILP model, which is solvable based on a cutting constraint method. [Wang et al. \(2013\)](#) developed two types of global optimization methods, and one of them involves using the system optimal (SO) traffic assignment principle to construct effective lower bounds. Recently, [Wang et al. \(2015\)](#) considered both which links should be built and how many capacities should be assigned to those built links simultaneously, and then proposed a global optimization method incorporating linearization, outer approximation and range reduction to solve this problem.

Those previous studies, based on either bi-level programming or transformed single-level models, have made great contributions to a deep understanding and numerical efficient algorithm development for different classes of discrete traffic network design problems. To the best of our knowledge, still very few studies have completely considered the tight transportation capacity and (reasonable and realistic) user equilibrium in the context of bi-level or single-level network design framework, especially for time-dependent transit service networks.

1.2 Transit service network design problem

This study takes particular interests in the dynamic transit service network design, as a representative example of discrete transportation service network design in a space-time network. There are a wide range of studies on transit network design, represented by excellent survey papers such as [Guihaire and Hao \(2008\)](#), [Kepaptsoglou and Karlaftis \(2009\)](#), [Farahani et al. \(2013\)](#), and [Ibarra-Rojas et al. \(2015\)](#). Table 1 lists a number of studies closely related to the problem under consideration in this paper, with a special focus on topics such as travelers' behavior in terms of

deterministic user equilibrium or Nash equilibrium, time-dependent travel demand, space-time network representation, and general transit assignment problems which is the lower level of the transit service network design problem.

Table 1. Related studies on user equilibrium, capacity constraint, or time-dependent travel demand in transit systems

Paper	Problem type	User equilibrium vs. System Optimal	Capacity constraint	Demand	Space-time network	Solution method
Nguyen and Pallottino (1988)	Transit assignment	Strategy-based UE represented by variational inequality	No	Static	No	Shortest hyperpath algorithm
Spieß and Florian (1989)	Transit assignment	Strategy-based UE represented by one convex cost differentiable problem	No	Static	No	Linear approximation based algorithm
De Cea and Fernandez (1993)	Transit assignment	Volume-delay function-based UE represented by variational inequality	No	Static	No	Diagonalization algorithm
Lam et al. (1999)	Transit assignment	Generalized Stochastic UE	Modified link cost function	Static	No	Lagrangian algorithm
Nguyen et al. (2001)	Transit assignment	Nash equilibrium represented by variational inequality	Modified link cost function	Time-dependent	No	Column generation scheme
Nuzzolo et al. (2001)	Transit assignment	Dynamic Stochastic user equilibrium	No	Responsive time-dependent	Yes	Iterative algorithm with network loading
Gao et al. (2004)	Transit network design	Volume-delay function-based UE represented by variational inequality	No	Static	No	Heuristic algorithm based on sensitivity analysis
Poon et al. (2004)	Transit assignment	Dynamic user equilibrium	Tight	Time-dependent	No	Iterative algorithm with network loading
Hamdouch et al. (2004)	Transit assignment	Strategy-based UE represented by variational inequality	Tight	Static	No	Iterative algorithm with network loading
Cepeda et al. (2006)	Transit assignment	Strategy-based UE represented by gap function	Modified link cost function	Static	No	Method of successive average (MSA)
Hamdouch and Lawphongpanich (2008)	Transit assignment	Strategy-based Dynamic UE represented by variational inequality	Tight	Time-dependent	Yes	Method of successive average (MSA) with network loading
Nuzzolo et al. (2012)	Transit assignment	Dynamic Stochastic user equilibrium	Tight	Responsive time-dependent	Yes	Iterative algorithm with network loading
Niu and Zhou (2013)	Transit service network design (Timetable optimization)	System-optimal	Tight	Time-dependent	Yes	Genetic algorithm
Hamdouch et al. (2014)	Transit assignment	Strategy-based Stochastic UE	Tight	Time-dependent	Yes	Iterative algorithm with network loading
Szeto and Jiang (2014a)	Transit assignment	Approach-based UE represented by variational inequality	Modified link cost function	Static	No	Extragradient method
Szeto and Jiang (2014b)	Transit network design	System-optimal, Bi-level model	Tight	Static	No	Hybrid artificial bee colony algorithm
Niu et al. (2015)	Transit service network design (Timetable optimization)	System-optimal	Tight	Time-dependent	Yes	Optimization solvers
Verbas and Mahmassani (2015)	Transit service network design (Frequency allocation)	Dynamic user equilibrium	Tight	Responsive time-dependent	No	Heuristic method and simulation via bi-level programming
Liu and Zhou (2016)	Transit service network design	Network equilibrium with boundedly rational agents	Tight	Time-dependent	Yes	Lagrangian decomposition

As listed in Table 1, in some early studies (Nguyen and Pallottino, 1988; Spiess and Florian, 1989; De Cea and Fernandez, 1993; Nuzzolo et al., 2001; Gao et al., 2004), the tight capacity constraints are not considered for user equilibrium in transit assignment/network design problems; in some succeeding studies, the tight capacity constraints are generally addressed in the following two ways: (i) extend or modify existing link travel time functions to penalize the generalized cost values when the assigned flow is above the capacity (Lam et al., 1999; Nguyen et al., 2001; Cepeda et al., 2006; Szeto and Jiang, 2014a), (ii) explicitly consider strict capacity constraints through an inequality where the assigned flow is strictly equal to or less than the given capacity, implemented by simulation-typed network loading for equilibrium condition (Poon et al., 2004; Hamdouch et al., 2004; Hamdouch and Lawphongpanich, 2008; Nuzzolo et al., 2012; Hamdouch et al., 2014; Verbas and Mahmassani, 2015) or analytically mathematic models for system optimum (Niu and Zhou, 2013; Szeto and Jiang, 2014b; Niu et al., 2015).

Further, to find one equilibrium condition in the schedule-based transit assignment problem with tight capacity constraints, most of previous studies adopted an iterative procedure with simulation-type network loading, where both the best path finding for dynamic travel demand and the network loading for path cost calculation are performed iteration by iteration until reaching convergence. Generally, there are three types of approaches to find the best paths, (i) the approach by Poon et al. (2004) finds the least-generalized cost path by the specific time-dependent optimal path algorithm (Tong and Richardson, 1984); (ii) Hamdouch and Lawphongpanich (2008) proposed the optimal strategy (a set of paths with least expected travel cost) for different passenger groups by dynamic programming; (iii) Nuzzolo et al. (2012) selected a set of paths with assignment probability by a route choice model while considering the real-time transit information. In this paper, the path selection mechanism we adopt is based on the boundedly rational travel behavior that the path cost of each agent should be within the sum of the agent's least path cost and indifference band. Since non-atomic game and atomic game belong to two different modeling frameworks, in order to avoid the confusion with the traditional BRUE with nonatomic players, the agent-based transit assignment result in this paper can be treated as a kind of network equilibrium with boundedly rational agents.

1.3 Wardrop's user equilibrium and bounded rationality behavior

A number of questions should be systematically addressed in order to analytically represent practically important modeling aspects, such as how to consider tight capacity constraints (e.g. vehicle carrying capacity and station spatial capacity) within UE-oriented behavioral assumptions. In an early study, Hearn (1980) stated that a real user equilibrium proposed by Wardrop may not exist in the traffic assignment problem with tight link capacity constraints. Adding the tight link capacity constraint to the Beckmann-McGuire-Winsten (BMW) model (Beckmann et al., 1956), Larsson and Patriksson (1995) proposed an augmented Lagrangean dual algorithm for a modified BMW model and showed that there exists a generalized UE where Lagrangean multipliers of link capacity constraints can be treated as link tolls. Further, Larsson and Patriksson (1995), Larsson and Patriksson (1998), Nie et al. (2004), and Marcotte et al. (2004) noted that those Lagrangian multipliers in generalized user equilibrium is not unique. In order to obtain a specific toll scheme (Lagrangian multipliers) for implementation, it usually requires modelers to consider a secondary goal (Larsson and Patriksson, 1998), such as, minimizing the total system tolling charges, minimizing the maximum toll on any individual arc, or minimizing the number of toll booths based on the tolling implementation conditions (Florian and Hearn, 1999). The other way to address tight capacity constraints is to use the marginal cost as link toll to achieve a generalized user equilibrium so that travelers can choose the routes (the solution) of system optimum (Florian and Hearn, 1999). In addition, the classical work by Correa et al. (2004) proves that the solution of the modified BMW model belongs to one of Nash equilibriums where no travelers can reduce their travel cost by unilaterally changing routes in a capacitated network.

Typically, once both the tight capacity and self-disutility minimization behavior (Wardrop's first principle) are considered, the problem can be modeled as a generalized Nash equilibrium problem (GNEP) (Facchinei and Kanzow, 2010). Each player has one disutility function and aims to choose a strategy that minimizes his/her disutility. Meanwhile, the player's strategy also depends on the rival players' strategies due to limited resources. However, it is still challenging to find a widely accepted solution approach to this complex problem, and more modeling details are systematically examined in the excellent survey paper by Facchinei and Kanzow (2010).

A theoretically important study by Szeto (2003) points out that dynamic user equilibrium (DUE) may not exist when considering the physical queue in dynamic traffic systems. Han et al. (2015a) first presented a rigorous continuity result for the path delay operator, which is the fundamental importance to the existence of DUE conditions, based on the Lighthill-Whitham-Richards (LWR) network model capable of capturing physical queues and spillback, by assuming that the network supply is bounded away from zero or is within desired boundedness. In addition, Han et al. (2015b) proposed a bi-level model for traffic network signal control problems where a continuum signal model is employed to ensure the existence of DUE in the lower level problem. As a remark, the dynamic user equilibrium studied in the last two decades is usually analytically expressed of Nash-like equilibrium condition (Han et al., 2015a).

In order to address the non-existence issue of flow-based dynamic user equilibrium, Szeto (2003) developed a tolerance-based dynamic traffic assignment (DTA) model, where the tolerance can also be treated as an indifference band, which can be viewed as an extension or adaption from the boundedly rational behavioral model first introduced through the seminar work by Mahmassani and Chang (1987). In those metropolitan's transit system under oversaturated conditions, the tight vehicle capacity constraints lead to possible discontinuity of path travel time and unfortunately force some travelers fail to board on the preferred line, which could result in travelers' boundedly rational behavior based on their day-to-day (possibly stochastic) travel experiences. Meanwhile, many empirical studies using GPS trajectory data (e.g., Morikawa et al., 2005; Zhu, 2011) have shown that travelers do not always choose the shortest paths in reality.

1.4 Potential contributions and structure of this paper

Motivated by the need for realistically modeling behavioral responses of each traveler to a wide range of emerging active demand management strategies, such as, personalized incentives, this paper aims to enhance the behavioral realism in the context of dynamic transit service network design under very tight capacity constraints. Different from traditional Wardrop equilibrium conditions that assume non-atomic players (i.e. infinitesimal number of travelers), we are actually considering a special network equilibrium with boundedly rational agents (NEBRA) through a single-level 0-1 integer linear programming formulation.

Specifically, (1) Focusing on a dynamic transit service network with tight capacity, we utilize the property of constant travel times on space-time arcs to formulate the boundedly rational travel behavior of each traveler through a set of integer linear inequalities; (2) By exogenously listing a set of viable space-time path alternatives for each agent, we offer a single-level 0-1 integer linear programming model to study the complex discrete network design problem under a set of quite specific but realistic assumptions. With the aim of minimizing the total transit system travel time, this new reformulation avoids the use of possible non-convex flow-based models where its link travel time is usually represented by a kind of nonlinear functions. (3) Although the proposed agent-based and time-dependent formulation introduces additional dimensions and a large number of binary variables, it is further shown that, after dualizing hard constraints (i.e., capacity and rational decision constraints), the original problem can be decomposed into two sub-problems. These sub-problems have computationally efficient algorithms available on large scale networks, namely a time-dependent least cost path problem and a knapsack problem.

The agent-based formulation and model decomposition scheme used in this paper can be viewed as further extensions of two recent papers (Tong et al., 2015; Li et al., 2015), which aim to maximize space-time accessibility and total system-wide utility. However, the main theme and contributions of this paper focus on how to theoretically model and reformulate the extremely critical user behavior under tight capacity restrictions on specific dynamic transit service network design problems.

The remainder of the paper is organized in the following manner. The following section illustrates the major features of our proposed NEBRA through a simple example. Section 3 formulates the static discrete network design problem with boundedly rational agents for the dynamic case presented in Section 4. Section 5 describes the Lagrangian relaxation framework for solving the proposed integer linear programming model. After discussing some practically important modelling features in Section 6, the performance of our proposed methodology is evaluated through several numerical experiments in Section 7.

2. Conceptual illustration

2.1 Notations

Table 2 lists general indices, sets, parameters and variables in optimization models appeared in this paper.

Table 2. Indices, sets, parameters and variables

Indices	Definition
i, j, j'	Index of nodes, $i, j, j' \in N$
(i, j)	Index of physical link between two adjacent nodes, $(i, j) \in L$
a	Index of agents, $a \in A$, defined based on each (time-dependent) OD pair
w	Index of OD pairs, $w \in W$
p	Index of path set P_w
k	Index of a path
$o(a)$	Index of origin node of agent a
$d(a)$	Index of destination node of agent a
t, s	Index of time intervals in the space-time network

Sets

N	Set of nodes in the physical transportation network
L	Set of links in the physical transportation network
W	Set of Origin-Destination (OD) pairs
A	Set of agents
P_w	Set of paths of OD pair w
L_C	Set of current links in the physical transportation network
L_B	Set of potential built links in the physical transportation network
$\Phi(w(a), k)$	Set of links in the k^{th} path of OD pair $w(a)$ of agent a
$\Omega(w(a))$	Set of possible paths of OD pair $w(a)$ of agent a
V	Set of vertices in the space-time network
E	Set of edges/arcs in the space-time network

Parameters

d^w	Travel demand of OD pair w
$\delta_{(i,j)}^{w,p}$	Path-link incidence index of route p of OD pair w on link (i, j)
$Cap_{i,j}$	Capacity of link (i, j)
$Cap_{i,j,t,s}$	Capacity of traveling arc (i, j, t, s) in the space-time network
$\varepsilon(a)$	The indifference band or tolerance value of agent a
ε_w	The indifference band of OD pair w
$b_{i,j}$	Construction cost of link (i, j)
B	Total financial budget
M	An assumed large value as an auxiliary parameter
DT^a	The departure time of agent a
AT^a	The assumed arrival time of agent a
$c_{i,j,t,s}$	Travel cost of traveling arc (i, j, t, s) in the space-time network
T	The time horizon in the space-time network
$BT_{w,DT}$	Budget time of all agents of OD pair w at departure time DT
$h_{w,p}$	Path flow of path p of OD pair w

Variables

$c_{i,j}$	Travel cost of link (i, j)
$\bar{c}_{i,j}$	Generalized travel cost of link (i, j)
C_w^a	Path cost of agent a which departs from OD pair w
π_w	The minimal path travel cost of OD pair w
$C_{w(a),p}$	Path cost of path p of OD pair w of agent a
$x_{i,j}^a$	$= 1$ if agent a is assigned on link (i, j) ; $= 0$ otherwise
$f_{i,j}$	Traveler flow on link (i, j)
$y_{i,j}$	$= 1$ if physical link (i, j) is decided to be constructed in the physical transportation network; $= 0$ otherwise
$x_{i,j,t,s}^a$	$= 1$, if Agent a is assigned on traveling/waiting arc (i, j, t, s) in the space-time network; $= 0$ otherwise
$y_{i,j,t,s}$	$= 1$ if service arc (i, j, t, s) is decided to be operated; $= 0$ otherwise

2.2 Network equilibrium condition with boundedly rational agents

Since Wardrop's user equilibrium may not exist due to strict capacity constraints, the gap of route travel time among travelers with same origin, destination, and departure time could force travelers to accept an indifference band through day-to-day travel experiences and finally form boundedly rational travel behavior. For each agent, his/her perceived travel cost on the selected route is constrained within the respective indifference bands given by:

$$C_w^a \leq \pi_w + \varepsilon(a), \forall a \quad (1)$$

It should be noted that, the traditional flow-based BRUE condition is described by the following "if-then" condition: if $h_{w,p} > 0$, then $C_{w,p} \leq \pi_w + \varepsilon_w$ (Lou et al., 2010; Guo and Liu, 2011; Di et al., 2013; Di et al., 2014; Han et al., 2015c; Di et al., 2016). In our proposed case, for any path selected by one or more agents, the corresponding path flow $h_{w,p}$ is equal to or more than 1. As a result, the path cost should be constrained within the predefined bound $\pi_w + \varepsilon_w$ for an OD pair, or equivalently $\pi_w + \varepsilon(a)$ which reflects an individual's indifference band.

For illustrative purposes, one simple network is created that was adapted from the paper by Correa et al. (2004), with 4 nodes and 6 links along one OD pair (1,4) shown in Fig.1. The link cost and link capacity are sequentially displayed in parenthesis. The demand from node 1 to node 4 has 2 agents.

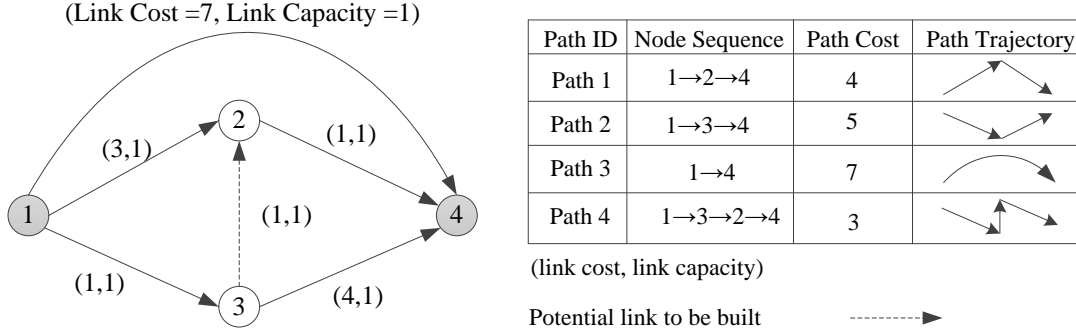


Fig. 1 One simple illustrative network modified from Correa et al. (2004)

The results are then examined under different traveling behavioral assumptions, including (i) generalized UE in the modified BMW model, (ii) system optimal, (iii) self-disutility minimization behavior (Wardrop's first principle), and (iv) NEBRA for our network design problem.

(i) The modified BMW model:

$$\min \sum_{(i,j) \in L} \int_0^{f_{i,j}} c_{i,j}(u) du \quad (2)$$

Subject to,

$$\sum_{p \in P_w} h_{w,p} = d_w, \forall w \in W \quad (3)$$

$$\sum_{w \in W} \sum_{p \in P_w} (\delta_{(i,j)}^{w,p} \times h_{w,p}) = f_{i,j}, \forall (i,j) \in L \quad (4)$$

$$f_{i,j} \leq Cap_{i,j}, \forall (i,j) \in L \quad (5)$$

$$h_{w,p} \geq 0, \forall p \in P_w, \forall w \in W \quad (6)$$

The formulation is the standard BMW model with adding tight link capacity constraints. Eq. (3) is the path flow conservation constraint. In Eq. (4) the link flow is equal to the path flow multiplied by the path-link incidence value. Inequality (5) is the tight link capacity constraint, and inequality (6) defines the path flow as a nonnegative continuous variable. As mentioned in Section 1, the generalized UE (GUE) includes the Lagrangean multipliers $\mu_{i,j}$ of link capacity constraints (5), and the GUE condition is derived from the KKT conditions shown in Appendix A. The generalized link cost can be represented as

$$\bar{c}_{i,j} = c_{i,j} + \mu_{i,j}, \forall (i,j) \in L \quad (7)$$

(ii) System optimal:

$$\min \sum_{(i,j) \in L} (c_{i,j}(f_{i,j}) \times f_{i,j}) \quad (8)$$

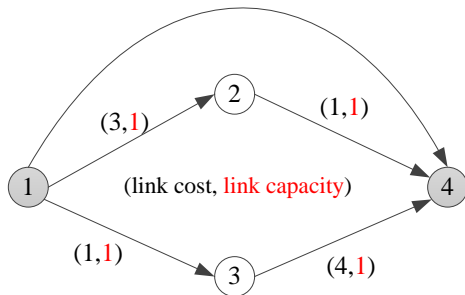
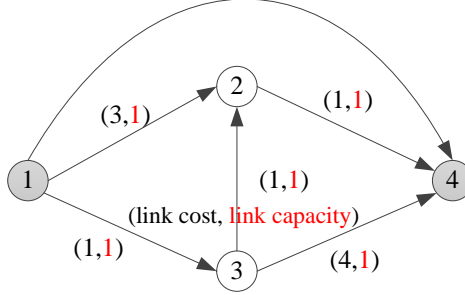







Subject to constraints (3)-(6).

(iii) Self-disutility minimization behavior (Wardrop's first principle): since it is a generalized Nash equilibrium problem and the model is difficult to solve, the solution is enumerated in this simple example.

(iv) NEBRA: the boundedly rational travel behavior has been formulated by inequality (1), and all possible solutions are also enumerated.

Table 3 lists comparison results for all cases.

Table 3. Comparison results of different cases for the network design problem

Cases	<div>(Link Cost =7, Link Capacity =1)</div> 						<div>(Link Cost =7, Link Capacity =1)</div> 						
	Without link 3 → 2						With link 3 → 2						
	Path 1 	Path 2 	Path 3 	Total Cost	π	Gap	Path 1 	Path 2 	Path 3 	Path 4 	Total Cost	π	Gap
1. Modified BMW	✓	✓	✗	$4 + 5 = 9$	4	1	✓	✓	✗	✗	$4 + 5 = 9$	3	3
2. System Optimal	✓	✓	✗	$4 + 5 = 9$	4	1	✓	✓	✗	✗	$4 + 5 = 9$	3	3
3. Self-disutility minimization	✓	✓	✗	$4 + 5 = 9$	4	1	✗	✗	✓	✓	$3 + 7 = 10$	3	4
4. NEBRA (0,0)	✗	✗	✗	N/A	4	N/A	✗	✗	✗	✗	N/A	3	N/A
5. NEBRA (0,1)	✓	✓	✗	$4 + 5 = 9$	4	1	✗	✗	✗	✗	N/A	3	N/A
6. NEBRA (0,4)	✓	✓	✗	$4 + 5 = 9$	4	1	✗	✗	✓	✓	$3 + 7 = 10$	3	4
	✓	✗	✓	$4 + 7 = 11$	4	3							
7. NEBRA (2,4)	✓	✓	✗	$4 + 5 = 9$	4	1	✓	✓	✗	✗	$4 + 5 = 9$	3	3
	✓	✗	✓	$4 + 7 = 11$	4	3	✗	✗	✓	✓	$3 + 7 = 10$	3	4
	✗	✓	✓	$5 + 7 = 12$	4	4	✓	✗	✓	✗	$4 + 7 = 11$	3	5
							✗	✓	✓	✗	$5 + 7 = 12$	3	6

NEBRA (0,1): one agent's indifference band value is 0, and the other one's is 1;

√: One agent will choose the corresponding path;

×: No agent will choose the corresponding path;

N/A: Not applicable. There is no feasible solution.

What we can observe from Table 3 is summarized as follows.

(1) Since the cost of all links is constant, the modified BMW model is the same as the system optimal model. In addition, no agent can reduce his/her travel cost by unilaterally changing routes, due to link capacity constraints in case 1 and case 2. This means that the two solutions belong to Nash equilibriums.

(2) In case 3, all travelers are assumed to be inclined to minimize their own disutility when selecting routes, which is consistent with Wardrop's first principle. Interestingly, one can identify the Braess paradox when building a new link. By enumeration, (i) if link $3 \rightarrow 2$ is not built, one agent will choose path 1 and the other will choose path 2, so the total system cost is 9 (money units); (ii) if link $3 \rightarrow 2$ is built, one agent will choose the new shortest path, path 4, and the other has to accept the only one available path, path 3, due to tight link capacity constraints, so the total system cost is 10.

(3) If link $3 \rightarrow 2$ is built, different solutions of cases 1, 2 and 3 satisfy the definition of Nash equilibrium, but the solution of case 3 is the only result obeying Wardrop's first principle in a capacitated network, that is, representing travelers' self-disutility minimization behavior with strict link capacity constraints. Further discussions along this line can be found in the paper by [Correa et al. \(2004\)](#). It is observed that the shortest path (path 4) is not used in cases 1 and 2, even the solution of the two cases meets the Nash equilibrium condition. It is consistent with the statement of [Larsson and Patriksson \(1995\)](#) about the result of generalized user equilibrium.

(4) The gap function we consider is $\sum_a (C_w^a - \pi_w)$. If link $3 \rightarrow 2$ is built, the gap for the first two cases is 3. With a larger gap value of 4, case 3 (applying Wardrop's first principle) does not lead to a smaller gap value in a capacitated network. The reason is that the gap function assumes that all travelers should be at the minimum cost paths after assignment ([Hearn, 1982](#)), but the tight capacity constraints cause some travelers have to accept a longer path in the final solution.

(5) Given the KKT condition (described in Appendix A) in case 1 with link $3 \rightarrow 2$, the Lagrange multipliers of links (3,2) and link (1,4) are 0 because the link flow is less than its link capacity. For the remaining links, those multipliers are not unique as long as the generalized path travel cost of path 1 and path 2 is equal and not more than that of the longest path (path 3). To some extent, this behavior is related to the argument of [Nie et al. \(2004\)](#) and [Marcotte et al. \(2004\)](#) that it may be questionable to treat those multipliers as queuing delays because of the non-uniqueness of the multiplier values.

(6) Now consider different cases for NEBRA with specific indifference bands for two individuals. In case 4, when the indifference band of two agents is 0, it is not surprising to see there is no solution for agents without tolerance. For a particular example where link $3 \rightarrow 2$ is not built, the specific set of inequalities for boundedly rational behavior can be listed as follows:

For agent 1: $C^1 \leq 4 + 0$

For agent 2: $C^2 \leq 4 + 0$

As shown in Fig.1, the path cost of the shortest path (path 1) is 4, but its capacity is only 1, so it is impossible for two agents to use one feasible path at the same time.

(7) In case 5, with the increased indifference band of 1 for the second agent, the feasible solution exists for the do-nothing case; but no NEBRA solution can be found if link $3 \rightarrow 2$ is built. In reality, when facing the change of network conditions, travelers may need to change their own indifference band in order to adapt to the new situations. Interested readers are referred to a recent paper by [Lo \(2013\)](#), which interprets similar behavior by highlighting that bounds on rationality are determined by physiological and environmental constraints. However, how to quantify the changing bounds on rationality is still a very complex research topic in its own right and beyond the scope of this paper.

(8) In case 6, (i) if link $3 \rightarrow 2$ is not built, there are two feasible solutions with different system cost values of 9 and 11, which can be viewed as the best case vs worst case of NEBRA solutions. (ii) if link $3 \rightarrow 2$ is built, there is only one feasible solution and the total cost is 10. As a result, if we consider the best case of NEBRA as the selected solution for the before and after scenarios, it would lead to the Braess paradox. However, the Braess paradox can be avoided in this particular case if the worst case of NEBRA is assumed. Therefore, selection of feasible solution(s) of NEBRA would affect the final network design decision, as are also clearly identified by a recent important paper by [Lou et al. \(2010\)](#). Additionally, in case 7, there are multiple solutions whether or not link $3 \rightarrow 2$ is built, so it is difficult to select a single solution from the before or after solution set to support the network design decision.

3. Modelling on a discrete network design problem with boundedly rational agents

In order to clearly illustrate the forthcoming dynamic transit service network design problem in Section 4, this section will focus on the static case as a starting point.

Consider a transportation network as a directed graph with N as the set of nodes and L as the set of links. Let W denote the set of OD pairs connected by the set of feasible paths, P_w . Each agent a is defined based on its origin $o(a)$ and destination $d(a)$ and has its own tolerance value $\varepsilon(a)$. Each link (i, j) has its travel cost $c_{i,j}$ and capacity $Cap_{i,j}$. In order to improve the total transportation system efficiency, a total financial budget B is planned and the construction cost of new line (i, j) is $b_{i,j}$.

The static discrete transportation network design with bounded-rational agents can be formulated as follows:

Objective function:

$$\min \sum_a \sum_{(i,j) \in L} (c_{i,j} \times x_{i,j}^a) \quad (9)$$

Subject to:

Budget constraint:

$$\sum_{(i,j) \in L} b_{i,j} \times y_{i,j} \leq B, \forall (i,j) \in L, b_{i,j} = 0 \text{ for } (i,j) \in L_c \quad (10)$$

Capacity constraint:

$$\sum_a x_{i,j}^a \leq y_{i,j} \times Cap_{i,j}, \forall (i,j) \in L, y_{i,j} = 1 \text{ for } (i,j) \in L_c \quad (11)$$

Flow balance constraint:

$$\sum_{i:(i,j) \in L} x_{i,j}^a - \sum_{i:(j,i) \in L} x_{j,i}^a = \begin{cases} -1 & \forall a, j = o(a) \\ 1 & \forall a, j = d(a) \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

Boundedly rational travel decision rule:

$$\sum_{(i,j) \in L} \{c_{i,j} \times x_{i,j}^a\} \leq \pi_{w(a)} + \varepsilon(a), \forall a \quad (13)$$

Definition of the shortest path:

$$\pi_{w(a)} = \min \{C_{w(a),1}, C_{w(a),2}, \dots, C_{w(a),k}\}, \forall k \in \Omega(w(a)) \quad (14)$$

Binary variables: $x_{i,j}^a = \{0,1\}$ and $y_{i,j} = \{0,1\}$

The objective function of this network design problem is to minimize the total transportation system travel cost of all agents. Inequality (10) represents the budget constraint, where the construction cost of existing links is 0. Inequality (11) is the link capacity constraint, where $y_{i,j} = 1$ for existing links. Eq. (12) is the standard agent-based flow balance constraint. Inequality (13) and Eq. (14) represent boundedly rational travel decision rule and the shortest path discussed above, respectively. They can be combined into the following set of inequalities (15):

$$\begin{cases} \sum_{(i,j) \in L} \{c_{i,j} \times x_{i,j}^a\} \leq C_{w(a),1} + \varepsilon(a) \\ \sum_{(i,j) \in L} \{c_{i,j} \times x_{i,j}^a\} \leq C_{w(a),2} + \varepsilon(a), \forall a \\ \dots \\ \sum_{(i,j) \in L} \{c_{i,j} \times x_{i,j}^a\} \leq C_{w(a),k} + \varepsilon(a) \end{cases} \quad (15)$$

Meanwhile, the path cost of feasible path k of OD pair w_a can be formulated as Eq. (16), where an auxiliary large value M is introduced to represent link cost as $c_{i,j} + (1 - y_{i,j}) \times M$. If link (i, j) is built, its cost is $c_{i,j}$; otherwise, its large cost value prevents any agent from selecting that link.

$$C_{w(a),k} = \sum_{(i,j) \in \Phi(w_a,k)} \{c_{i,j} + (1 - y_{i,j}) \times M\} \quad (16)$$

Therefore, the boundedly rational travel decision rule can be reformulated as follows:

$$\sum_{(i,j) \in L} \{c_{i,j} \times x_{i,j}^a\} \leq \sum_{(i,j) \in \Phi(w_a,k)} \{c_{i,j} + (1 - y_{i,j}) \times M\} + \varepsilon(a), \forall a, \forall k \in \Omega(w(a)) \quad (17)$$

A simple study case is created for illustrating the transformation above in Fig. 2. One agent (agent a) departs from origin node 1 to destination 5 with indifference band of 1 time unit. Link (3,5) and link (4,5) are potentially built.

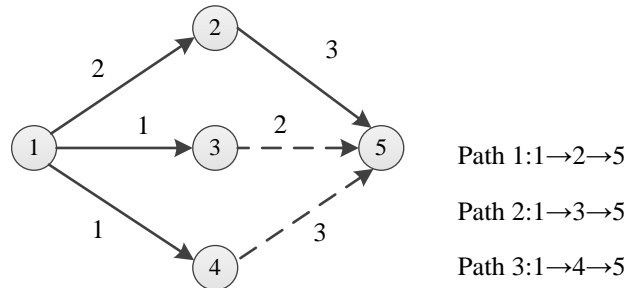


Fig. 2 Simple study case for the illustration above

The boundedly rational travel decision rule can be represented as

$$\sum_{(i,j) \in L} \{c_{i,j} \times x_{i,j}^a\} \leq C_{w(a),1} + \varepsilon(a) = 5 + 1 \quad (17.a)$$

$$\sum_{(i,j) \in L} \{c_{i,j} \times x_{i,j}^a\} \leq C_{w(a),2} + \varepsilon(a) = 1 + 2 + (1 - y_{3,5}) \times M + 1 \quad (17.b)$$

$$\sum_{(i,j) \in L} \{c_{i,j} \times x_{i,j}^a\} \leq C_{w(a),3} + \varepsilon(a) = 1 + 3 + (1 - y_{4,5}) \times M + 1 \quad (17.c)$$

The possible chosen path set of agent a is listed in Table 4. The different assumed network design decisions could lead to different shortest paths for one specific OD pair, so the tightest and active constraint of boundedly rational travel decision rule could be different, which bounds the feasible path set to each agent.

Table 4. Possible chosen path set of agent a under different network design decisions

Network design decision	The least path cost π	Rational bound $\pi + \varepsilon(a)$	The tightest constraint of boundedly rational travel decision rule	Feasible path set of agent a
$y_{3,5} = 0, y_{4,5} = 0$	5	6	Inequality (17.a)	Path 1
$y_{3,5} = 1, y_{4,5} = 0$	3	4	Inequality (17.b)	Path 2
$y_{3,5} = 0, y_{4,5} = 1$	4	5	Inequality (17.c)	Path 3 and path 1
$y_{3,5} = 1, y_{4,5} = 1$	3	4	Inequality (17.b)	Path 2 and path 3

As a result, the critical variable π_w for the boundedly rational travel decision rule does not appear in the model and it is implicitly defined through a set of path cost inequalities (17). This requires to exogenously enumerate a number of paths for the OD pair of each agent, including the shortest path in the existing network and those paths that have potential built links and are shorter than the shortest path in the existing network. Typical route set generation algorithms can be found in the dissertation by [Ramming \(2002\)](#) for large-scale networks. For simplicity, the K-shortest path algorithm ([Yen, 1971](#); [Xu, et al., 2012](#)) can also be adopted to seek the potential shortest path π_w for each OD pair.

In the typical bi-level programming structure, the lower level UE problem only considers the links to be built decided from the upper level problem (and existing links). In our proposed single-level model, the boundedly rational travel decision rule is represented by considering the complete set of all possible paths that embed the decision variable $y_{i,j}$ covering two cases: to be built or not to be built in the final optimal solution. More specifically, if $y_{i,j} = 0$ for the link not to be built, the right-hand-side path cost is infinity, allowing inequality (17) to hold in any case. If $y_{i,j} = 1$ for the link to be built in the final solution, (i) capacity inequality (11) allows the corresponding link flow to be positive; (ii) a feasible path in the final solution will have a path cost $C_{w,p} < \infty$. Furthermore, if there is a feasible solution for all agents' boundedly rational behavior with $\pi_w + \varepsilon(a) < \infty$, then each agent must be able to select a capacity-feasible path p from the path set $\Omega(w(a))$ satisfying $C_{w,p} \leq \pi_w + \varepsilon(a) < \infty$.

The non-uniqueness of traditional BRUE solutions has been studied or discussed in several papers ([Lou et al., 2010](#); [Guo and Liu, 2011](#); [Di et al., 2013](#); [Di et al., 2014](#); [Han et al., 2015c](#); [Di et al., 2016](#)). More specifically, most researchers focus on the best case (risk-averse), the worst case (risk-prone), and the neutral case (risk-neutral) for the unique solution of BRUE. In this paper, since the objective function of our proposed single-level programming model (presented herein), $\min_{y,x(y)} f(y,x)$, is to minimize the total travel time, it indicates that the best case of the NEBRA solution is assumed as travelers' responses. Usually, $\min_y \min_{x(y)} f(y,x)$ is one representation for the objective function of network design problems with the best case of traditional BRUE conditions ([Di et al., 2016](#)) from the bi-level programming perspective. In order to handle a very complex relationship between x and y represented through the lower level optimization problem in the bi-level formulation, the most straightforward way to find the optimal solution is to enumerate all feasible y , solve the corresponding lower level model $\min_{x(y)} f(y,x)$, and then compare all $\min_{x(y)} f(y,x)$ to find the minimum of $\min_{x(y)} f(y,x)$, which is $\min_y \min_{x(y)} f(y,x)$. It is important to note that, within the same solution space of x and y , in our proposed single-level programming model, the relation among y and x has been clearly defined in constraints, and the objective function $\min_{y,x(y)} f(y,x)$ can also be viewed to find the minimum of $\min_{x(y)} f(y,x)$ based on all possible y . Therefore, we can state that only the best case is considered in our model. In the future, we will conduct further studies to examine different assumptions about travelers' boundedly rational decision behavior in our agent-based framework.

Typically, the discussions on the existence and uniqueness of traditional flow-BRUE are built on nonlinear and convex programming techniques, while the proposed NEBRA formulation in this paper is an integer linear programming model in nature. In many cases, there might be multiple path solutions even for the simplest single OD pair shortest path problem in a grid network. Thus, the properties (solution existence and uniqueness) of our proposed model will be dependent on one specific integer linear programming model on different time-discretized networks, and the further discussions of those solution properties are beyond the scope of this paper.

4. Dynamic discrete transit service network design with boundedly rational agents

4.1 Construct the transit service space-time network

In order to represent (i) the transit service network based on the given train/bus schedule, (ii) transit vehicle capacity, and (iii) transit station/platform storage capacity, we extend the physical transit network into a space-time network from the given timetable. Consider a physical transit network with a set of nodes (stations) N and a set of links L . Each link can be denoted as a directed link (i, j) from upstream node i to downstream node j , with one deterministic scheduled travel time. We then construct a space-time network, where V is the set of vertices and E is the set of edges/arcs. Node i is extended to a set of vertices (i, t) at each time interval t in the study horizon, $t = 1, 2, \dots, T$. Each agent a , where $a \in A$, is assumed to have a planned departure time DT^a at origin node $o(a)$ to its destination node $d(a)$. At each destination node, there is one assumed large arrival time T for all agents. Meanwhile, set the travel cost of waiting arcs on the destination node as 0 in the space-time network to represent the end-to-end trip time. There are two following types of arcs.

(1) Link traveling arcs are extended from a link (i, j) and each arc traverses from vertex (i, t) to vertex (j, s) based on the given timetable, where $(s - t)$ is the scheduled link travel time and should be integer multipliers of one time interval and capacity is defined as the transit vehicle capacity.

(2) Waiting arcs from (i, t) to $(i, t + 1)$ at any node i have waiting time as one time interval and capacity is defined as the station/platform storage capacity. The waiting cost is set as 0 for the destination stations.

In addition, to address the common line issue in transit systems that there are multiple same schedules (travelling arcs) directly joining two vertexes in the space-time network, we can separate one vertex as multiple vertexes, each of which corresponds to one transit schedule. That method is similar to the approach used by [Poon et al. \(2004\)](#) and [Hamdouch and Lawphongpanich \(2008\)](#).

In order to find what impacts travelers' route choice most, one very interesting travel survey was performed in the Chicago Metropolitan area in 2010 ([Nie et al., 2010](#)), which reveals that over 80% of the responders choose travel time as their most important concern in their route choice. The remaining factors sorted by importance are reliability, cost, comfort and convenience, safety, and emission and energy conversation, respectively. Therefore, our first modeling priority regarding the general travel cost is still on travel time, including in-vehicle travel time, at-station/stop waiting time and transfer time. [Nuzzolo et al. \(2001\)](#) and [Nuzzolo et al. \(2012\)](#) considered a flow-based path travel cost composed of travel time (in-vehicle time, waiting time and transfer time), number of transfers, and in-vehicle discomfort cost. [Hamdouch and Lawphongpanich \(2008\)](#) considered travel time, transit fare, penalty for early/late arrival and early departure, and in-vehicle discomfort cost. In those papers, discomfort cost is represented by a flow-based link cost function, and all other factors are weighted by different given parameters. With our proposed space-time network structure, one can further incorporate various weights of in-vehicle travel time, waiting time, and arc-based transit fare as constant costs of different types of arcs. The modelling issue related to OD-based dynamic transit fare and penalty for early/late arrival, as part of general path cost, will be discussed in section 6.1 and 6.3, respectively.

For the illustrative purposes, Fig. 3(a) depicts a simple transit physical network with three nodes and two links. The demand has three agents departing at time 0 from node 1 to node 3, whose tolerance values/indifference bands are 0, 2, and 2, respectively. The station capacity at each node is assumed to be 2. A potential open new line is from node 1 to node 3. Table 5 lists the vehicle capacity parameters and scheduled travel time, corresponding to the space-time network presented in Fig. 3(b).

Table 5. Vehicle capacity and its scheduled travel time

Transit line with its vehicle		Departure Time	Vehicle Capacity	Schedule Travel Time		
				Link (1,2)	Link (2,3)	Link (1,3)
Existing Line(1 \rightarrow 2 \rightarrow 3)	Vehicle 1	0	1	1	2	\times
	Vehicle 2	2	2	1	2	\times
Potential Line(1 \rightarrow 3)	Vehicle 3	1	1	\times	\times	3

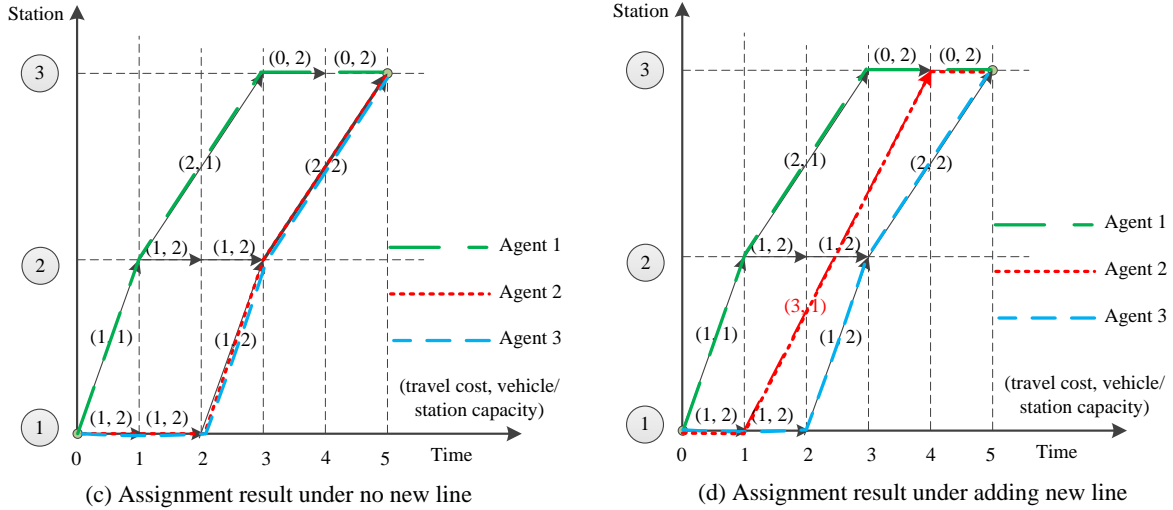
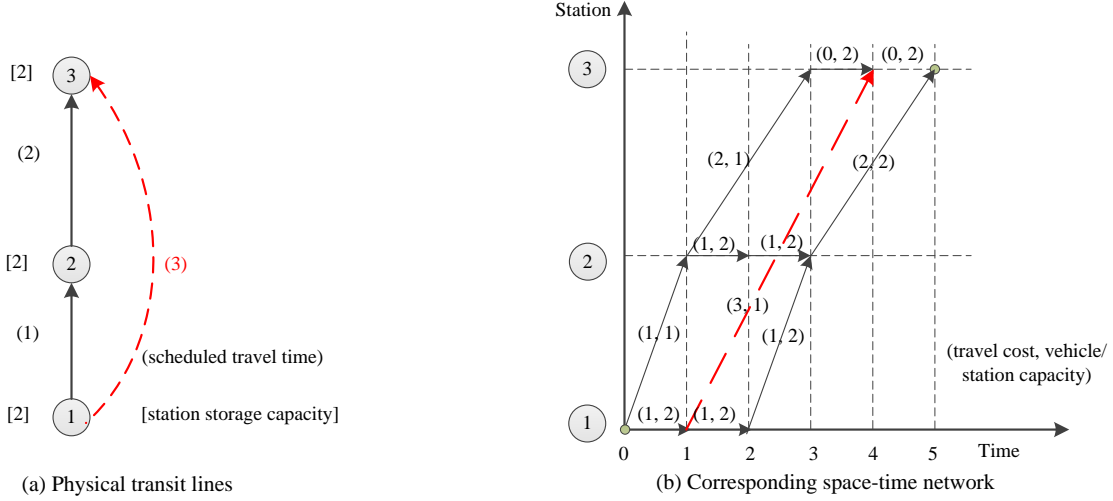


Fig. 3 Illustration of physical transit network, corresponding space-time network and assignment results

In the space-time network, the origin becomes the vertex $(1,0)$ and the destination is the vertex $(3,5)$, which is a single origin to single destination problem (one-to-one network). Based on the tolerance values of each agent and boundedly rational travel decision rule, $C_w^a \leq \pi_w + \varepsilon(a)$, the set of inequalities for each agent can be written as: $\pi_w = 3$, $\varepsilon(1) = 0$, $\varepsilon(2) = 2$, $\varepsilon(3) = 2$, so $C^1 \leq 3 + 0$, $C^2 \leq 3 + 2$, $C^3 \leq 3 + 2$. If new transit line $(1,3)$ is not built, the single solution satisfying the set of inequalities is shown in Fig. 3(c). Agent 1 will choose path $(1,0) \rightarrow (2,1) \rightarrow (3,3)$. Due to the vehicle capacity constraint, both agent 2 and agent 3 have to wait at station 1 for the next available vehicle at time 2, so the selected path will be $(1,0) \rightarrow (1,1) \rightarrow (1,2) \rightarrow (2,3) \rightarrow (3,5)$ with a total system-wide travel time of 13. After the new transit line is open, the best case solution is that agents 1 and 3 still choose the previous path and agent 2 will choose path $(1,0) \rightarrow (1,1) \rightarrow (3,4) \rightarrow (3,5)$, shown in Fig. 3(d) with the total system-wide travel time as 12. Table 6 provides a more systematic comparison between different scenarios and it should be noted that the agent number is only used for representing one traveler. It doesn't mean that agent 1 has the priority to choose a shorter path compared with other agents.

Table 6. Assignment result comparison of scenarios with and without adding new line

Scenarios	Agent ID	Agent 1	Agent 2	Agent 3
	Tolerance band	0	2	2
No new line	Inequality of boundedly rational travel decision rule	$C^1 \leq 3 + 0$	$C^2 \leq 3 + 2$	$C^3 \leq 3 + 2$
	Path node sequence	$(1,0) \rightarrow (2,1) \rightarrow (3,3)$	$(1,0) \rightarrow (1,1) \rightarrow (1,2) \rightarrow (2,3) \rightarrow (3,5)$	$(1,0) \rightarrow (1,1) \rightarrow (1,2) \rightarrow (2,3) \rightarrow (3,5)$

	Total cost	13		
Adding new line	Inequality of boundedly rational travel decision rule	$C^1 \leq 3 + 0$	$C^2 \leq 3 + 2$	$C^3 \leq 3 + 2$
	Path node sequence	$(1,0) \rightarrow (2,1) \rightarrow (3,3)$	$(1,0) \rightarrow (1,1) \rightarrow (3,4) \rightarrow (3,5)$	$(1,0) \rightarrow (1,1) \rightarrow (1,2) \rightarrow (2,3) \rightarrow (3,5)$
	Total cost	12		

When addressing the many-origin-to-many-destination networks, dummy node is required as the centroid for those destination nodes. More specific details about the extended space-time network with dummy nodes can be found in a major early paper by [Drissi-Kaitouni and Hameda-Benchekroun \(1992\)](#).

4.2 Proposed integer linear programming model

Prior to introducing the proposed model, the following key assumptions are presented as careful response to many potential modeling issues discussed in the aforementioned illustrative examples.

(1) The travel cost of each agent just considers the travel time on travelling arcs and waiting arcs on the extended space-time network.

(2) The time-dependent transit demand is given and deterministic.

(3) The tolerance value of each agent is given and does not change with respect to changes in the transit service network.

(4) Regarding the possible multiple solutions of NEBRA, the best case is chosen as the selected solution.

(5) For simplicity, the first-in-first-out (FIFO) rule is not considered in these dynamic transit systems, because the non-FIFO phenomenon still exists under certain conditions. For example, when travelers are waiting at a platform for the next desirable transit vehicle, it may not actually be the next available vehicle. Also, when travelers transfer to another station or board a transit vehicle, it is possible for them not to entirely obey the FIFO rule in reality.

The following presents the objective function and constraints applied in the proposed model.

Objective function:

$$\min \sum_a \sum_{(i,j,t,s) \in E} (c_{i,j,t,s} \times x_{i,j,t,s}^a) \quad (18)$$

Subject to,

Budget constraint:

$$\sum_{(i,j) \in L} b_{i,j} \times y_{i,j} \leq B, \forall (i,j) \in L, b_{i,j} = 0 \text{ for } (i,j) \in L_C \quad (19)$$

Capacity constraint:

$$\sum_a x_{i,j,t,s}^a \leq y_{i,j} \times \text{Cap}_{i,j,t,s}, \forall (i,j) \in L, y_{i,j} = 1 \text{ for } (i,j) \in L_C \quad (20)$$

Flow balance constraint:

$$\sum_{i,t:(i,j,t,s) \in E} x_{i,j,t,s}^a - \sum_{i,t:(j,i,s,t) \in E} x_{j,i,s,t}^a = \begin{cases} -1 & \forall a, j = o(a), s = DT^a \\ 1 & \forall a, j = d(a), s = T \\ 0 & \text{otherwise} \end{cases} \quad (21)$$

Boundedly rational travel decision rule:

$$\sum_{(i,j,t,s) \in E} \{c_{i,j,t,s} \times x_{i,j,t,s}^a\} \leq \sum_{(i,j,t,s) \in \Phi(w_a,k)} \{c_{i,j,t,s} + (1 - y_{i,j}) \times M\} + \varepsilon(a), \forall a, \forall k \in \Omega(w(a)) \quad (22)$$

Binary variables:

$$x_{i,j,t,s}^a = \{0,1\} \quad (23)$$

$$y_{i,j} = \{0,1\} \quad (24)$$

Similar to the objective function and side constraints of the static case presented in Section 3, this formulation is extended for the dynamic case in the space-time network based on the given schedule of existing transit lines and potential built or open service lines. Each agent chooses a set of arcs in the space-time network, so the decision variable $x_{i,j,t,s}^a$ is a binary variable. We need to recognize that once those binary variables on each arc are required to be aggregated for flow-based arc/link attributes, such as, for in-vehicle discomfort cost function, it will be a challenge to solve this problem. This is because the objective function (18) becomes a nonlinear function with binary variables. However, if the agent is treated as a group of passengers with the same attributes (such as, same indifference band) and the variable $x_{i,j,t,s}^a$ is further relaxed as a continuous variable rather than an integer variable, the problem will become a solvable flow-based BRUE network design problem capable of considering different link cost functions. In addition, $y_{i,j}$ in constraints (19), (20) and (24) could also be extended as $y_{i,j,t,s}$ when considering whether or not to add or close specific service arcs defined by the time-dependent schedule in the space-time network. In short, after problem decomposition in Section 5, the unique feature of our agent-based formulation with constant arc costs allows

us to handle much simpler time-dependent shortest path subproblems, compared to general multi-commodity flow-based formulations.

5. Lagrangian decomposition based solution procedure

5.1 Problem decomposition

Within a commonly used Lagrangian relaxation framework, a set of nonnegative capacity constraint multipliers $\mu_{i,j,t,s}$ and tolerance bound constraint multipliers λ_k^a are defined to dualize capacity constraints (20) and boundedly rational travel decision rule (22), respectively, onto the objective function (18) presented in Section 4 to generate lower bounds. The objective function is now transformed to

$$\min \sum_a \sum_{(i,j,t,s) \in E} (c_{i,j,t,s} \times x_{i,j,t,s}^a) + \sum_{(i,j,t,s) \in E} [\mu_{i,j,t,s} \times (\sum_a x_{i,j,t,s}^a - y_{i,j} \times Cap_{i,j,t,s})] + \sum_a \sum_k \lambda_k^a \times \{\sum_{(i,j,t,s) \in E} (c_{i,j,t,s} \times x_{i,j,t,s}^a) - \sum_{(i,j,t,s) \in \Phi(w_a,k)} [c_{i,j,t,s} + (1 - y_{i,j}) \times M] - \varepsilon(a)\} \quad (25)$$

Subject to constraints (19), (21), (23), and (24).

Based on the decision variables $x_{i,j,t,s}^a$ and $y_{i,j}$, the dualized problem above can be decomposed into two subproblems P_x and P_y .

Subproblem P_x for finding time-dependent shortest path for each agent:

$$\min \sum_a \sum_{(i,j,t,s) \in E} (c_{i,j,t,s} + \mu_{i,j,t,s} + \sum_k \lambda_k^a \times c_{i,j,t,s}) \times x_{i,j,t,s}^a \quad (26)$$

Subject to (21) and (23).

Subproblem P_y as a binary knapsack problem for service arc selections:

$$\max \sum_{(i,j,t,s) \in E} \mu_{i,j,t,s} \times y_{i,j} \times Cap_{i,j,t,s} + \sum_a \sum_k \lambda_k^a \times \{\sum_{(i,j,t,s) \in \Phi(w_a,k)} [c_{i,j,t,s} + (1 - y_{i,j}) \times M] + \varepsilon(a)\} \quad (27)$$

Subject to (19) and (24).

5.2 Lagrangian relaxation based algorithm design

The general procedure of the Lagrangian relaxation-based algorithm is designed as shown in Fig. 4, and one can apply dynamic transit simulation to find upper bound feasible solutions based on the network design decisions from the lower bound model (i.e. the relaxed problem).

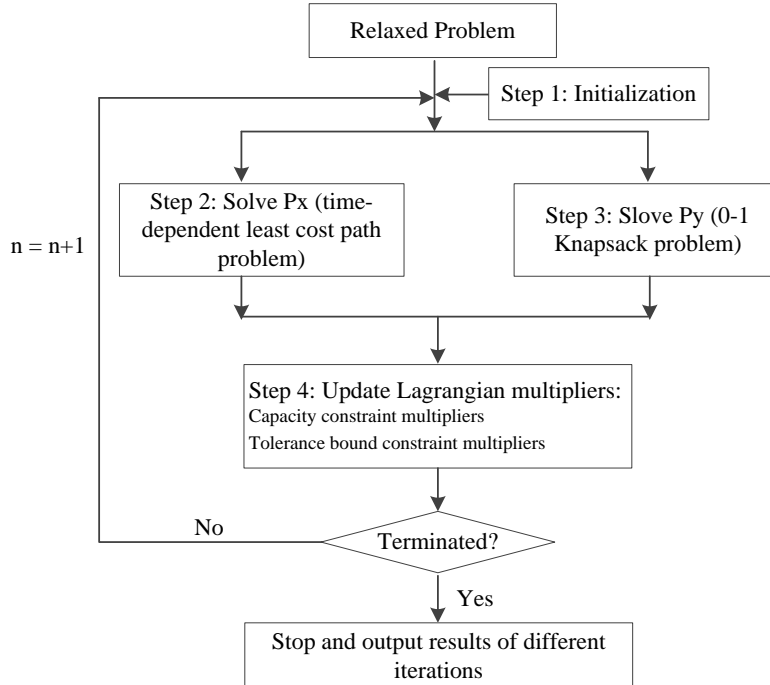


Fig. 4 Flowchart of solving the relaxed problem under Lagrangian relaxation framework

The specific procedure is described as follows:

Step 1: Initialization

Initialize iteration number $n = 0$;

Construct the space-time network based on the given transit physical network and time schedule;

1 Initialize the set of Lagrangian multipliers $\mu_{i,j,t,s}$ and λ_k^a as positive values.
2 Step 2: Solve subproblem P_x for $x_{i,j,t,s}^a$.
3 Step 3: Solve subproblem P_y for $y_{i,j}$.
4 Step 4: Update Lagrangian multipliers:
5 4.1 Calculate the subgradients:
6 Subgradient of capacity constraint: $\nabla\mu_{i,j,t,s} = \sum_a x_{i,j,t,s}^a - y_{i,j} \times Cap_{i,j,t,s}$
7 Subgradient of tolerance bound constraint: $\nabla\lambda_k^a = \sum_{(i,j,t,s) \in E} (c_{i,j,t,s} \times x_{i,j,t,s}^a) -$
8 $\sum_{(i,j,t,s) \in \Phi(w_a,k)} [c_{i,j,t,s} + (1 - y_{i,j}) \times M] - \varepsilon(a)$
9 4.2 Update Lagrangian multipliers:
10 Multiplier of capacity constraint: $\mu_{i,j,t,s}^{n+1} = \max\{0, \mu_{i,j,t,s}^n + \alpha^n \times \nabla\mu_{i,j,t,s}\}$
11 Multiplier of tolerance bound constraint: $\lambda_k^a(n+1) = \max\{0, \lambda_k^a(n) + \alpha^n \times \nabla\lambda_k^a\}$
12 Where, α^n is the step size, and $\alpha^n = 1/(n+1)$.
13 Step 5: Termination condition test
14 If n is equal to the predetermined maximum iteration number N , terminate the algorithm; otherwise,
15 $n = n + 1$ and go back to Step 2 with updated $\mu_{i,j,t,s}$ and λ_k^a values.
16 At step 2, all arcs have constant travel cost in the space-time network, and the Lagrangian multipliers are given
17 as constant values at each iteration. The subproblem P_x becomes a standard time-dependent least cost path problem.
18 At step 3, the arc travel cost, arc capacity, Lagrangian multipliers, big M , and tolerance value of each agent are given
19 and constant. The subproblem P_y becomes a standard 0-1 knapsack problem. [Ziliaskopoulos and Mahmassani \(1993\)](#)
20 and [Pallottino and Scutellà \(1998\)](#) provided more details about the space-time network construction and time-
21 dependent least cost path finding algorithms, which does not require the addition of a dummy node for the
22 corresponding destination node in many-origin-to-many-destination networks. For the 0-1 knapsack problem, there
23 are several available computationally efficient algorithms based on dynamic programming, branch and bound, or the
24 hybridizations of the both ([Martello and Toth, 1990](#)).
25

6. Discussions

6.1 Transit pricing

27 Transit pricing is one important factor affecting final decisions on the transit service network design. To consider
28 time-dependent and personalized agent-based or OD based (given) pricing parameters, one can extend the space time
29 arc cost $c_{i,j,t,s}$ to $c_{i,j,t,s}^a$, and add dummy starting arcs to represent OD-specific or agent-specific price for related travel
30 distance or/and traveling time periods. That is, for each OD or agent, there is one dummy link with a specific cost
31 value. Considering the price or incentive as part of the decision variables in the network design problem, different
32 pricing/incentive strategies can be listed using different dummy arcs. Then add a restriction that only one of the dummy
33 arc sets for each time-dependent OD pair or agents can be selected for the final strategy to be used in the final solution.
34
35

6.2 Passenger inflow rate control to transit stations

36 One approach for addressing the safety concern in a totally oversaturated condition is to limit the number of
37 passengers in the station. This practice was implemented in actual subway operations during peak hours in Beijing
38 ([Xu et al., 2014](#)). To model this inflow passenger volume gating strategy, a virtual node needs to be added to the
39 related origin node in the physical network. A simple network with two nodes and one link is used to illustrate our
40 method as shown in Fig. 5(a). There are two vehicles with capacity of 5 and scheduled travel time of 3, departing at
41 time 1 and 3, respectively. The station/platform capacity of node 1 and 2 is 10. Now assume that the permitted inflow
42 rate is 7. The change in the physical network is shown in Fig. 5(a). Link(1', 1) is treated as the entry point to the
43 station and its capacity represents the permitted passenger inflow rate. The corresponding space-time network is shown
44 in Fig. 5(b). If the number of agents at node 1' at time 0 is more than 7 at first, some agents will need to wait at node
45 1' until time 2 to enter the station for the second vehicle.
46

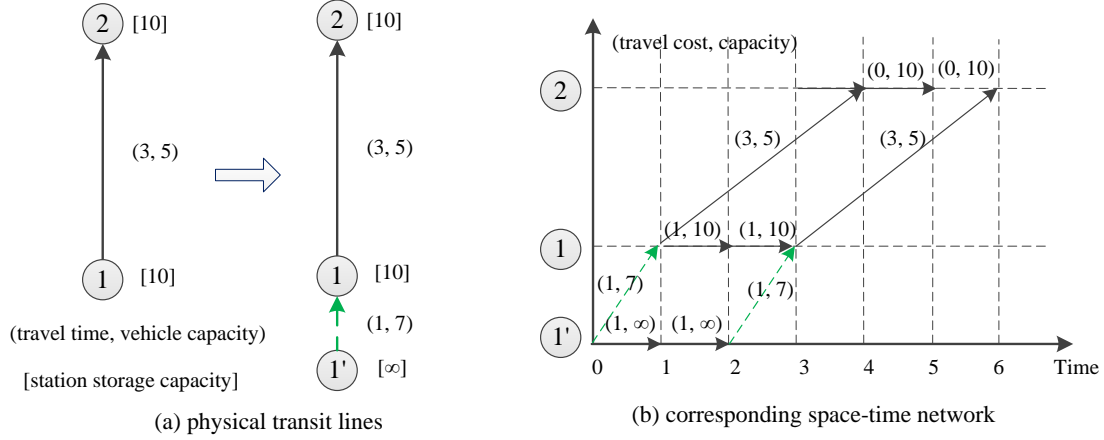


Fig. 5 original and modified space-time networks for modeling

6.3 Model the penalty for early/late arrival

In addition to the indifferent/tolerance value, each traveler may have one specific preferred arrival time for his or her trip within the bound. The penalty for early and later arrival had been considered in some traffic and transit studies (such as, Zhou et al., 2008; Hamdouch and Lawphongpanich, 2008). In this section, the issue is considered through updating the space-time network, where each destination node will be correspondingly given a virtual node as a super destination. The travelling arcs from the real destination node to the virtual node can be viewed as penalty arcs, the cost of which is defined in advance for each agent as part of his or her own path travel cost.

The network without walk arcs in the paper (Hamdouch and Lawphongpanich, 2008) is chosen as our illustrative example. Based on the given schedule of three lines, the corresponding space-time network is built in Fig. 6. Assume that agent 1 departs from node 1 to node 4 at time 1 with preferred arrival time of 5 and agent 2 has the same OD pair of agent 1 with different departure time and preferred arrival time, which are time 3 and time 7, respectively. The penalties $c_{i,j,t,s}^a$ for the early and late arrival of agent 1 is predefined as $c_{4,4',4,11}^1 = 1$, $c_{4,4',5,11}^1 = 0$, $c_{4,4',7,11}^1 = 4$, $c_{4,4',9,11}^1 = 8$ and $c_{4,4',10,11}^1 = 10$. Also, the early or late arrival penalties of agent 2 can be defined in advance in the space-time network. As a result, the arc cost in the space-time network should be represented by $c_{i,j,t,s}^a$ instead of $c_{i,j,t,s}$ in the objective function and the boundedly rational travel decision rule. Except for those penalty arcs, the remaining arc cost of each agent $c_{i,j,t,s}^a$ is still equal to $c_{i,j,t,s}$. Also, it is applicable by changing every arc cost if users impose different weights or parameters on different types of arcs, such as, in-vehicle travel arcs, waiting and transfer arcs, and penalty arcs for early and late arrival.

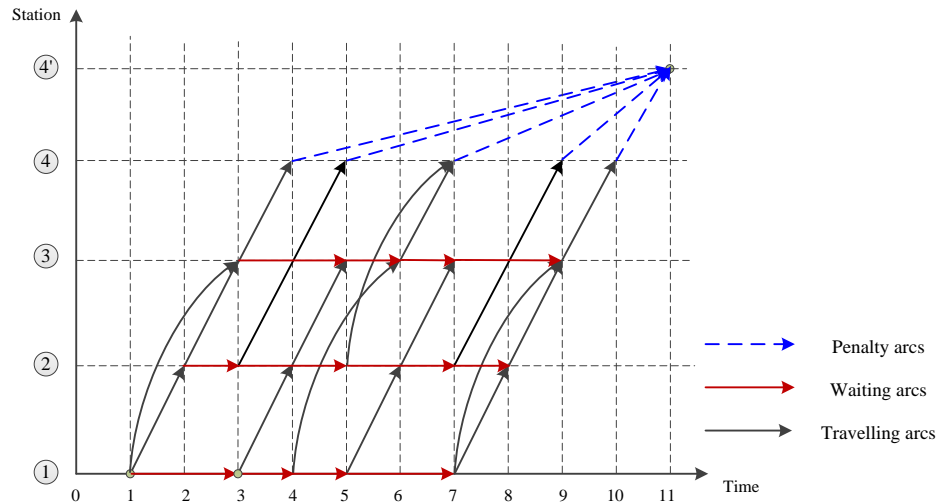


Fig. 6 Modified space-time network with early and later arrival penalty arcs

6.4 Adding travel time budget to simplify path enumeration on large-scale transit networks

The model proposed in section 4.2 requires to enumerate all possible paths for each agent. For a large-scale transit network, it will be extremely burdensome to perform the space-time path enumeration task. For simplification, the acceptable bound for each agent shown in inequality (22) can be assumed to be a constant travel time budget $BT_{w,DT}^a$ based on each OD pair at different departure times. In this approach, only a limited number of paths satisfying the following constraint are required.

$$\sum_{(i,j,t,s) \in E} \{c_{i,j,t,s} \times x_{i,j,t,s}^a\} \leq BT_{w,DT}^a \quad (28)$$

Usually, the travel time budget can be obtained through a travel survey. A real-world survey including travel time budget was performed in Chicago Metropolitan area (Nie et al., 2010). The detailed discussion can be found in this review paper (Mokhtarian and Chen, 2004), which states that individuals' travel time expenditures do show patterns and are strongly related to individual and household characteristics, attributes of activities at the destination, and characteristics of residential areas.

In addition, constraint (28) can be further modeled as virtual arcs to represent possible inaccessibility for this OD pair under congested condition. As a result, inequality (28) will be eliminated in our model and is simply coded in the space-time network as an approximation. The related application can be found at accessibility-based network design problem (Tong et al., 2015). Meanwhile, this virtual arc-based network modeling method could greatly reduce the computational complexity for large-scale applications, because it does not need to consider a set of constraints (constraint 28) in the primal problem or a set of dualized constraints in the Lagrangian relaxation procedure.

7. Numerical examples

This section examines different aspects in a dynamic discrete transit service network design problem, where (i) the scheduled travel time is constant and (ii) the capacity on travelling arcs and waiting arcs can represent the vehicle capacity and the station/platform capacity, respectively. The integer linear programming model and the proposed Lagrangian relaxation procedure are demonstrated by the general purpose optimization package GAMS (Rosenthal, 2015) in two small transit networks, and further be tested by our time-dependent shortest path algorithm in C++, by enhancing an open-source mesoscopic dynamic traffic assignment model namely DTALite (Zhou and Taylor, 2014), for large-scale applications on one workstation with 20 CPU cores (40 threads) with 192GB RAM. The related source codes can be downloaded at the website:

https://www.researchgate.net/publication/295968354_Case_1_Transit_network_design_LR
and https://www.researchgate.net/publication/295968184_Case_2_Demand_Level_1_Transit_network_design_LR

7.1. A simple case

This case examines the simple transit network in section 4.1 with an opening cost of 10 units and a total budget of 15 units for the potential open line. All Lagrangian multipliers are initially assumed to be 0.1, and the big $M = 6$ is sufficient because the time horizon is just 5. In order to not enumerate all possible paths of OD pair (1,3), at first, the shortest path is found in the existing space-time network and then all possible paths that contain the potential open lines are compared. As a result, $k = 2$ for OD pair (1,3). Path 1 is $(1,2,0,1) \rightarrow (2,3,1,3) \rightarrow (3,3,3,4) \rightarrow (3,3,4,5)$ and path 2 is $(1,1,0,1) \rightarrow (1,3,1,4) \rightarrow (3,3,4,5)$.

Based on the procedure of Lagrangian relaxation proposed in section 5.2, the objective values of P_x , P_y , and the lower bound are obtained and listed in Table 7 after 10 iterations.

Table 7. The objective values of P_x , P_y , and the lower bound during 10 iterations

Iterations	1	2	3	4	5	6	7	8	9	10
P_x	12	15.5	18.5	21.93	21.45	22.07	21.95	22.36	22.04	22.35
P_y	7.7	8.9	10.9	10.47	9.72	10.32	10.08	10.51	10.14	10.47
Z_{lb}	4.3	6.6	7.6	11.47	11.73	11.75	11.87	11.85	11.90	11.88

Fig. 8 demonstrates a comparison between the lower bound and the optimal value with a final solution gap of 0.83%.

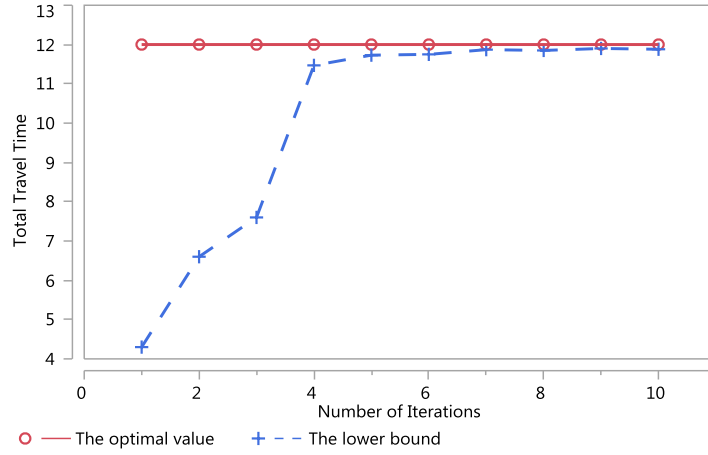


Fig.8 The comparison between the lower bound and the optimal

7.2. Transit service network design based on the simplified Sioux-Fall physical network

The proposed algorithm will be also tested in the following network with four hypothetical transit lines, shown in Fig. 9.

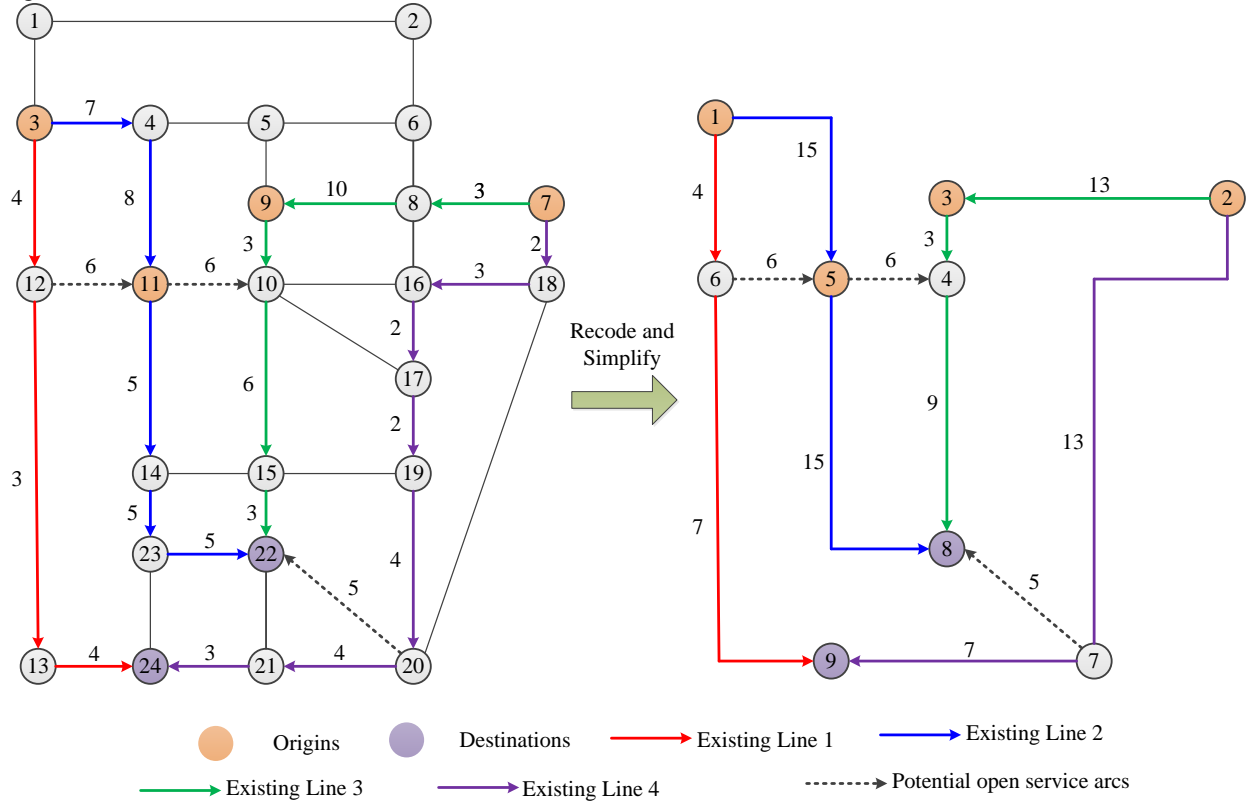


Fig. 9 Hypothetic Sioux-Fall transit network

Table 8 lists the existing transit service arcs based on given the timetable of the four transit lines. The potential service arcs to be opened and related station storage capacity are listed in Table 9 and Table 10, respectively. All serviced arcs are displayed in the space-time network shown in Fig. 10, where the virtual waiting cost at the destination 8 and 9 is 0.

Table 8. Hypothetic existing transit service arcs

Service Arc	Start Time	End Time	Service Arc	Start Time	End Time
(1,6)	0	4	(4,8)	16	25

(6,9)	4	11	(2,3)	15	28
(1,6)	15	19	(3,4)	28	31
(6,9)	19	26	(4,8)	31	40
(1,6)	30	34	(2,3)	30	43
(6,9)	34	41	(3,4)	43	46
(1,5)	0	15	(4,8)	46	55
(5,8)	15	30	(2,7)	0	13
(1,5)	15	30	(7,9)	13	20
(5,8)	30	45	(2,7)	15	28
(1,5)	30	45	(7,9)	28	35
(5,8)	45	60	(2,7)	30	43
(2,3)	0	13	(7,9)	43	50
(3,4)	13	16			

Table 9. Potential open transit service arcs

Service Arc	Start Time	End Time	Service Arc	Start Time	End Time	Service Arc	Start Time	End Time
(6,5)	4	10	(5,4)	10	16	(7,8)	13	18
(6,5)	19	25	(5,4)	25	31	(7,8)	28	33
(6,5)	34	40	(5,4)	40	46	(7,8)	43	48

Table 10. Station storage capacity

Station	1	2	3	4	5	6	7	8	9
storage capacity	40	40	30	30	30	30	30	M	M

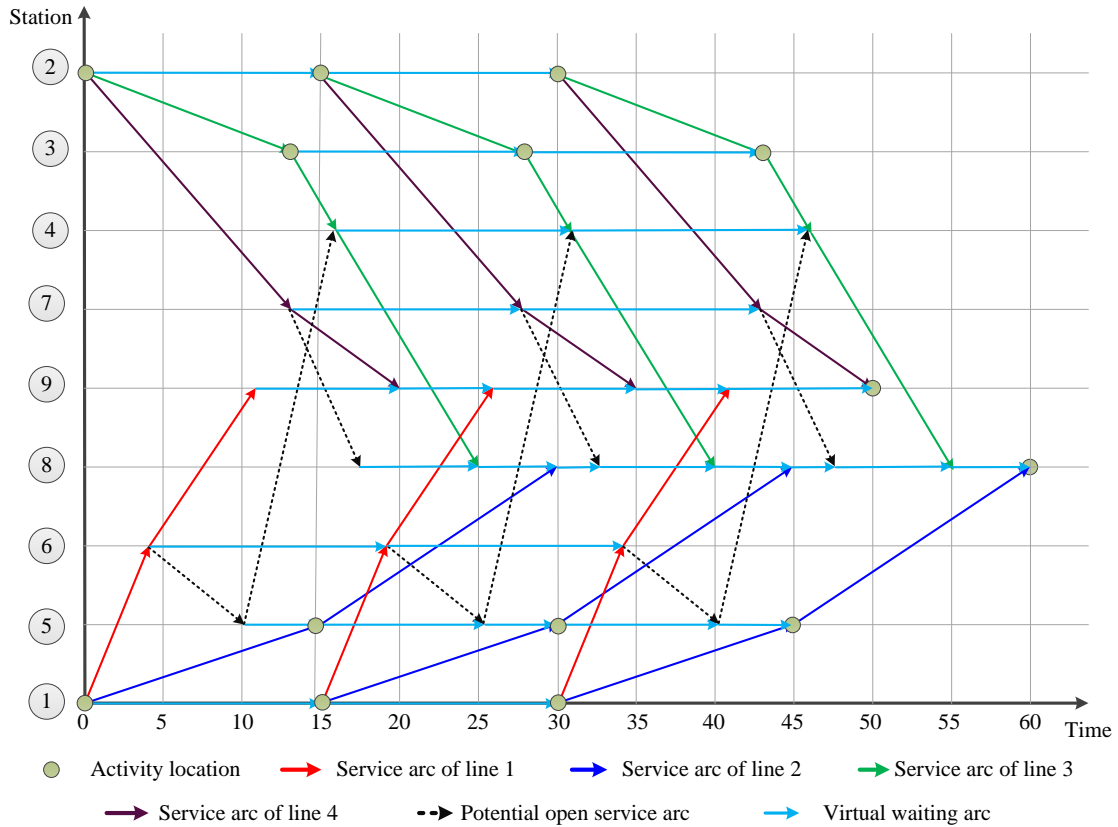


Fig. 10 The corresponding space-time transit service network

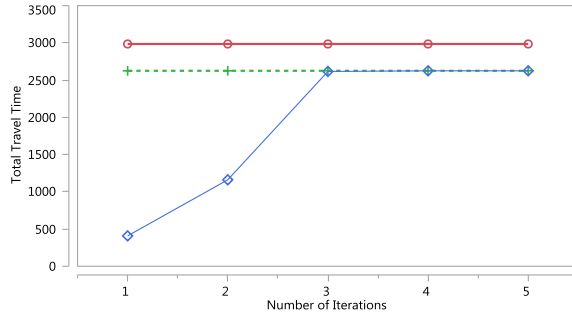
7.2.1 Different levels of time-dependent transit OD demand

The time-dependent OD demand for all OD pairs is listed in Table 11. Initially, three levels of demand are provided to observe its impact on the final service network design result. In addition, all vehicles' capacity is assumed to be 20. For simplicity, the tolerance value of all agents is assumed to be same as 15 (min or any time unit).

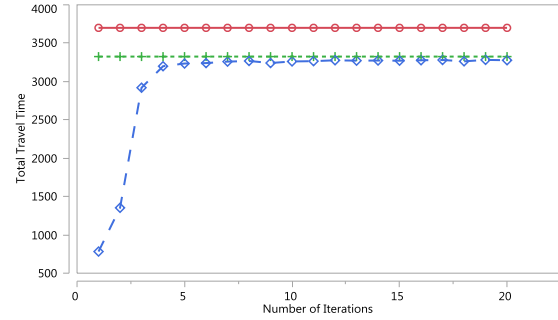
Table 11. Time-dependent OD demand

OD	Departure Time 1	Demand (agents)			Departure Time 2	Demand(agents)			Departure Time 3	Demand(agents)		
		L1	L2	L3		L1	L2	L3		L1	L2	L3
1→8	0	10	15	15	15	10	10	10	30	10	10	10
1→9	0	10	15	20	15	10	10	15	30	10	10	10
5→8	15	5	10	10	30	5	5	5	45	5	5	5
2→8	0	10	15	15	15	10	10	10	30	10	10	10
2→9	0	10	15	20	15	10	10	15	30	10	10	10
3→8	13	5	10	10	28	5	5	5	43	5	5	5

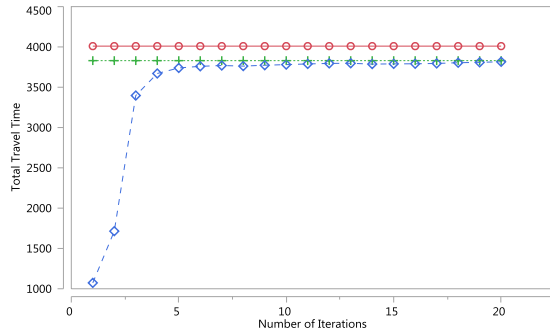
For each demand level, three cases are compared: (i) based on the existing transit service network, solved by the dynamic transit assignment problem with bounded rational user equilibrium conditions by GAMS; (ii) considering with potential new service arcs, optimal solution directly solved by GAMS for our proposed model in section 4.2; (iii) with potential new service arcs, solved using the proposed Lagrangian relaxation method implemented in GAMS. Comparison of the three cases under three demand levels are shown in Fig. 11 (a), Fig. 11(b) and Fig. 11(c), respectively.



(a) Demand level 1



(b) Demand level 2



(c) Demand level 3

Fig. 11 Comparison of the system-wide travel time of the three cases under three demand levels

Result presented in Fig.11 indicates that the total system-wide travel time gets reduced after transit service network optimization and our proposed Lagrangian relaxation procedure is able to achieve a quick convergence to the optimal solution.

In addition, in order to check the computation efficiency of GAMS solver and the proposed Lagrangian relaxation method, two high level OD demand inputs (demand level 4 and 5) are designed and presented in Table 12.

Table 12. Two high level time-dependent OD demand

OD		Demand(agents)		Demand(agents)		Demand(agents)
----	--	----------------	--	----------------	--	----------------

	Departure time 1	L4	L5	Departure time 2	L4	L5	Departure time 3	L4	L5
1→8	0	60	75	15	40	50	30	40	50
1→9	0	80	100	15	60	75	30	40	50
5→8	15	40	50	30	20	25	45	20	25
2→8	0	60	75	15	40	50	30	40	50
2→9	0	80	100	15	60	75	30	40	50
3→8	13	40	50	28	20	25	43	20	25

The comparison result on computation efficiency under five different levels of OD demand is displayed in Fig. 12. The total number of agents of demand level 1 to 5 is 150, 180, 200, 800, and 1000, respectively. It is obvious that the computation CPU time of the GAMS solver for optimal solutions is significantly increased under demand level 4 and 5, up to more than 40 mins. In comparison, the computation CPU time of the proposed Lagrangian relaxation method always remains low for all 5 demand levels considered. Even under demand level 5, its computation CPU time is less than 3 mins through 50 iterations, about only 6% of that required by the GAMS solver.

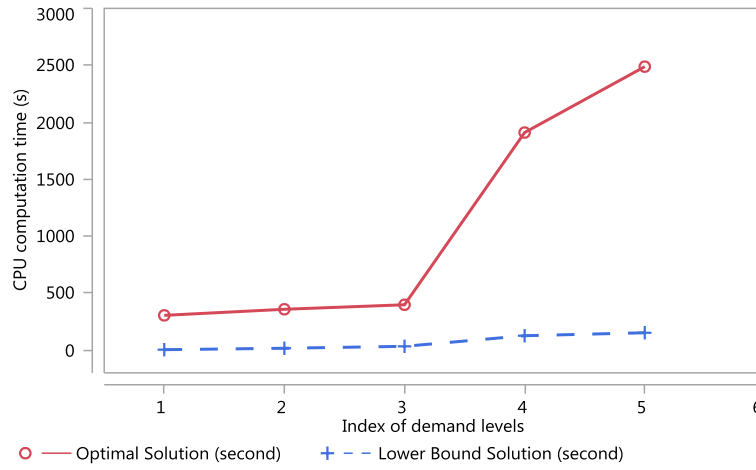


Fig. 12 Comparison on CPU computation time under different demand levels

7.2.2 Considering different levels of transit vehicle capacity

In order to perform the sensitivity analysis on the vehicle capacity, demand level 2 is adopted as the time-dependent OD demand input. The different vehicle capacities considered are 18, 19, 20, 21, and 22, respectively. The optimal system-wide travel time under different cases are shown in Fig. 13. It shows that based on the optimal solutions in the network design problem, the transit system efficiency improves with increased capacity of all vehicles. That is rational in that the larger vehicle capacity increase the service capacity of the whole network, including the shortest paths among OD pairs.

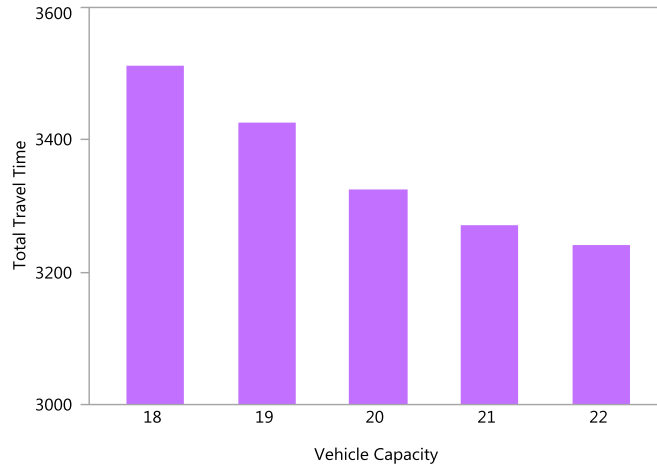


Fig. 13 Comparison on system-wide travel time under different vehicle capacities

7.2.3 Different levels of agents' tolerance value

The agent's tolerance value can also influence the final optimal solutions, so this section examines six scenarios to observe their impacts on the transit service network design. Scenario 1 to 4 have the same tolerance value for all agents, which is 7, 8, 9, and 10, respectively. For scenario 5, different agents could have different specific tolerance values as shown in Table 13. For example, for OD pair (1→8), there are 5 agents departing at time 0 with tolerance value of 0, 5 agents with tolerance value of 5, and 5 agents with tolerance value of 10.

Table 13. Number of agents with a specific tolerance value for scenario 5

OD	Departure time 1	Tolerance value			Departure time 2	Tolerance value			Departure time 3	Tolerance value		
		0	5	10		0	5	10		0	5	10
1→8	0	5	5	5	15	5	0	5	30	5	0	5
1→9	0	5	5	5	15	5	0	5	30	5	0	5
5→8	15	5	0	5	30	0	0	5	45	0	0	5
2→8	0	5	5	5	15	5	0	5	30	5	0	5
2→9	0	5	5	5	15	5	0	5	30	5	0	5
3→8	13	5	0	5	28	0	0	5	43	0	0	5

For scenario 6, the boundedly rational travel decision rule is ignored to obtain one system optimal solution for the transit service network design problem. The system performance results directly solved in GAMS are listed in Table 14.

Table 14. System performance of different scenarios

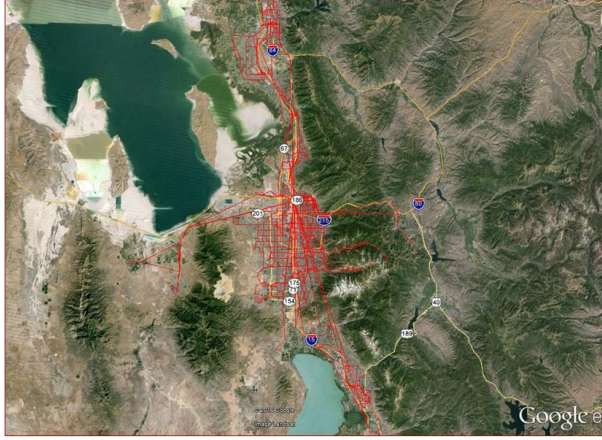
Scenario No.	Tolerance value	Total system travel time
1	7	Infeasible
2	8	3500
3	9	3500
4	10	3325
5	Specified	3400
6	System Optimal	3325

The results yield the following observations:

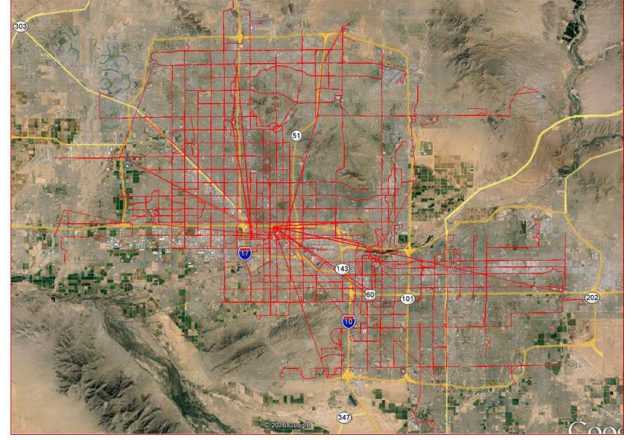
- (1) If all agents' tolerance value is 7, no feasible solution exists, which means that some agents will cancel their trips or the tolerance value is not accurate and needs further calibration
- (2) As the tolerance values of all agents increase, the system cost is closer to the system optimal value, because more agents can tolerance a longer path to allow some agents to select a shorter path
- (3) When the specific tolerance value of each agent is different, the transit system will have a more complicated route choice set.

7.3 Large-scale experiments

In order to demonstrate the computational efficiency of our proposed Lagrangian relaxation framework, this section tests our proposed algorithm on two regional transit networks, Salt Lake City regional transit network and Phoenix regional transit network, respectively. The transit network data are based on the public GTFS (**General Transit Feed Specification**) feed. Through extracting the transit stop information in *stops.txt* and stop sequences of each trip in *stop_times.txt*, the two transit networks are created and then are visualized in Google Earth as shown in Fig. 14. It can be observed that there are some unrealistically long straight transit links in the two networks. The reason is that the geographical shape information of each transit line in *shapes.txt* is not read. We just utilize the longitude and latitude coordinates of each stop of each transit line in *stops.txt*, and then one straight link is correspondingly created to connect the two joint stops.



(a) Salt Lake City regional transit network visualized in Google Earth



(b) Phoenix regional transit network visualized in Google Earth

Fig. 14 Salt Lake City and Phoenix regional transit feed data visualized in Google Earth, respectively

In order to obtain the dynamic transit travel OD demand data, two scenarios are assumed for the tests, (i) 10% of regional traffic demands as the transit OD demand input and (ii) 5% of regional traffic demands as the transit OD demand input. The details are listed in Table 15.

Table 15. Transit travel OD demand input of two scenarios

Networks	Demand time period	Total traffic demand	Transit demand (10%)	Transit demand (5%)
Salt Lake City	15pm-18pm	1.35 million agents	135,000 agents	67,500 agents
Phoenix	all day (24 hours)	10.5 million agents	1.05 million agents	525,000 agents

In addition, the activity location of each zone in the regional traffic network is matched with the nearest transit stop, so the final transit network information is listed in Table 16.

Table 16. Transit network information

Transit networks	Number of stops	Number of links connecting two joint stops at each transit line	Number of zones
Salt Lake City	6393	7219	2302
Phoenix	6788	7015	3022

Based on the simplification method in section 6.4 and the Lagrangian relaxation framework in Section 5, the relaxed problem can be decomposed as two subproblems at each iteration: one is time-dependent least cost path problem for each agent, which is the most time-consuming part in computation, and the other is the knapsack problem for service arc selection. In our proposed algorithm implemented in C++ code, Open Multi-Processing (OpenMP) is used as the application programming interface (API) for parallel computing. The tasks of path finding for each agent on the transit network are assigned into different available CPU threads and performed in parallel. The two regional transit networks under two different scenarios are tested on our Dell Precision T7610 Workstation with 20 CPU cores and 192G RAM. The general CPU computation time of each iteration in the Lagrangian relaxation approach are listed in Table 17.

Table 17. CPU computation time of one iteration in the Lagrangian relaxation approach

Transit networks	Scenario 1: Transit demand (10%)	Scenario 2: Transit demand (5%)
Salt Lake City	2 min 11 seconds	1 min 22 seconds
Phoenix	12 min 30 seconds	6 min 45 seconds

The CPU computation time could be affected by the number of optimized agents, the number of zones, the optimized time period, the topology of transit networks, the time budget value, etc. Looking back upon the study cases solved in GAMS in section 7.2, our proposed algorithm coded by C++ with the parallel computing method has an obvious advantage at computational efficiency. In future, based on the service arc selection decisions from the lower bound, the dynamic transit simulation with bounded rational agents will be developed for the upper bound. Finally, the gap between the lower bound and the upper bound can be used to check the quality of our solutions.

8. Conclusions and future research

The developments presented in this paper provide a unified framework for solving a dynamic transit service network design problem in congested networks where the service capacity and travelers' bounded rationality are highly respected. Within a space-time network, the time-dependent transit schedules and station services are represented by travelling arcs and waiting arcs with constant travel times. A single-level integer linear programming model is proposed that considers tight capacity constraints and the boundedly rational traveler behavior assumption to better capture the essential characteristics of transit dynamics. Computation efficiency is further enhanced by using Lagrangian relaxation approach which decomposes the primal problem as two sub-problems, namely, a time-dependent least cost path problem and a 0-1 knapsack problem. The numerical experiments described herein demonstrate the benefits of the proposed methodology and analyze the impacts of different transit demand levels, agent's tolerance value/indifference band, and transit vehicle capacity on the final service network decisions.

The transportation network design problem is usually modeled as a bi-level programming problem. The proposed framework from modelling to algorithm design is also helpful to problems requiring the similar bi-level programming structure, such as, dynamic tolling design problem, signal optimization problem, and dynamic OD demand matrix estimation problem. Future work includes (i) analyzing the properties of the NEBRA, especially in a general case where the arc cost in the space-time network could be linear or nonlinear functions, and calibrate the boundedly rational behavior by taking additional data collection efforts, (ii) considering the limited number of transfers through adding one more transfer state dimension, which can be referred to the time-dependent state-dependent dynamic programming algorithm for the pickup and delivery problem with time windows (PDPTW) by Mahmoudi and Zhou (2016), and (iii) addressing the responsive travel demand issue under different transit system performances. In order to implement more deployable service network design systems, the future study needs to incorporate realistic transit route/mode/departure choice models, e.g., by specifically taking into account of parking availability (Ruan et al., 2016), stochastic transfer activities (Yang et al., 2016), and time-dependent waiting time due to limited vehicle capacity (Yin et al., 2016).

In addition, the three path selection methods discussed in section 1.2 (by Poon et al., 2004; Hamdouch and Lawphongpanich, 2008; Nuzzolo et al., 2012) and our proposed NEBRA-based method has not been compared in this paper, as there is a potential inconsistency of input data and traveler behavior assumptions. For example, the travel behavior on route choice in different models has different characteristics. To name a few, the experienced least-cost path selection in method (i), optimal strategy for expected least-cost path set selection involving probabilistic behavior assumption in method (ii), path selection with perceived random errors in method (iii) and path selection with bounded rationality for each agent in our proposed method. Future research will be extremely important if one could use a rich set of real world data (with impacts from tight transit capacities) to more carefully examine, calibrate and validate the above models.

Appendix A: the KKT condition of the modified BMW model

$$\sum_{(i,j) \in L} (c_{i,j} + \delta_{(i,j)}^{w,p} \times \mu_{i,j}) \geq \pi_w, \forall p \in P_w, \forall (i,j) \in L \quad (\text{A.1})$$

$$\mu_{i,j} \geq 0, \forall (i,j) \in L \quad (\text{A.2})$$

$$\sum_{p \in P_w} h_{w,p} = d_w, \forall w \in W \quad (\text{A.3})$$

$$h_{w,p} \geq 0, \forall p \in P_w, \forall w \in W \quad (\text{A.4})$$

$$\sum_{w \in W} \sum_{p \in P_w} (\delta_{(i,j)}^{w,p} \times h_{w,p}) \leq \text{Cap}_{i,j}, \forall (i,j) \in L \quad (\text{A.5})$$

$$h_{w,p} (c_{i,j} + \delta_{(i,j)}^{w,p} \times \mu_{i,j} - \pi_w) = 0, \forall p \in P_w, \forall w \in W, \forall (i,j) \in L \quad (\text{A.6})$$

$$\mu_{i,j} (\text{Cap}_{i,j} - \sum_{w \in W} \sum_{p \in P_w} (\delta_{(i,j)}^{w,p} \times h_{w,p})) = 0, \forall (i,j) \in L \quad (\text{A.7})$$

Acknowledgements

We appreciate many insightful comments from a number of scholars in this field, especially Dr. Hani S. Mahmassani at Northwestern University and Dr. Jeffrey Stempihar at Arizona State University. We also appreciate three anonymous reviewers for their constructive comments. This paper is partially supported by National Science Foundation – United States under Grant No. CMMI 1538105 “Collaborative Research: Improving Spatial Observability of Dynamic Traffic Systems through Active Mobile Sensor Networks and Crowdsourced Data”. The work presented in this paper remains the sole responsibility of the authors.

References

- 1 Bard, J., Moore, J., 1990. A branch and bound algorithm for the bilevel programming problem. *SIAM Journal on*
2 *Scientific and Statistical Computing* 11 (2), 281–292.
- 3 Beckmann, M. J., McGuire, C. B., Winsten, C. B., 1956. *Studies in the Economics of Transportation*. Yale University
4 Press, New Haven, CT.
- 5 Cepeda, M., Cominetti, R., Florian, M., 2006. A frequency-based assignment model for congested transit networks
6 with strict capacity constraints: characterization and computation of equilibrium. *Transportation Research Part B*
7 40 (6), 437–459.
- 8 Correa, J. R., Schulz, A. S., Stier Moses, N. E., 2004. Selfish routing in capacitated networks. *Mathematics of*
9 *Operations Research* 29(4), 961–976.
- 10 Di, X., Liu, H.X., Pang, J.S., Ban, X., 2013. Boundedly rational user equilibria (BRUE): mathematical formulation
11 and solution sets. *Transportation Research Part B* 57, 300–313.
- 12 Di, X., He, X., Guo, X., Liu, H.X., 2014. Braess paradox under the boundedly rational user equilibria. *Transportation*
13 *Research Part B* 67, 86–108.
- 14 Di, X., Liu, H.X., Ban, X., 2016. Second best toll pricing within the framework of bounded rationality. *Transportation*
15 *Research Part B* 83, 74–90.
- 16 Drezner, Z., Wesolowsky, G.O., 2003. Network design: selection and design of links and facility location.
17 *Transportation Research Part A* 37 (3), 241–256.
- 18 Drissi-Kaïtouni, O., Hameda-Benchekroun, A., 1992. A dynamic traffic assignment model and a solution algorithm.
19 *Transportation Science* 26, 119–128.
- 20 Facchinei, F., Kanzow, C., 2010. Generalized Nash equilibrium problems. *Annual of Operations Research*. 175, 177–
21 211.
- 22 Farahani, R.Z., Miandoabchi, E., Szeto, W., Rashidi, H., 2013. A review of urban transportation network design
23 problems. *European Journal of Operational Research* 229 (2), 281–302.
- 24 Farvaresh, H., Sepehri, M.M., 2011. A single-level mixed integer linear formulation for a bi-level discrete network
25 design problem. *Transportation Research Part E* 47 (5), 623–640.
- 26 Florian, M., Hearn, D., 1999. Network equilibrium and pricing. In *Handbook of Transportation Science*. Springer US,
27 361–393.
- 28 Gao, Z., Sun, H., Shan, L., 2003. A continuous equilibrium network design model and algorithm for transit systems.
29 *Transportation Research Part B* 38, 235–250.
- 30 Gao Z, Wu J, Sun H, 2005. Solution algorithm for the bilevel discrete network design problem. *Transportation*
31 *Research Part B* 39, 479–95.
- 32 Guihaire, V., Hao, J.K., 2008. Transit network design and scheduling: A global review. *Transportation Research Part*
33 *A* 42 (10), 1251–1273.
- 34 Guo, X., Liu, H., 2011. Bounded rationality and irreversible network change. *Transportation Research Part B* 45 (10),
35 1606–1618.
- 36 Hamdouch, Y., Marcotte, P., Nguyen, S., 2004. Capacitated transit assignment with loading priorities. *Mathematical*
37 *Programming* 101B, 205–230.
- 38 Hamdouch, Y., Lawphongpanich, S., 2008. Schedule-based transit assignment model with travel strategies and
39 capacity constraints. *Transportation Research Part B* 42, 663–684.
- 40 Hamdouch, Y., Szeto, W.Y., and Jiang, Y., 2014. A New Schedule-Based Transit Assignment Model with Travel
41 Strategies and Supply Uncertainties. *Transportation Research Part B* 67, 35–67.
- 42 Han, K., Szeto, W.Y., Friesz, T.L., 2015a. Formulation, existence, and computation of boundedly rational dynamic
43 user equilibrium with fixed or endogenous user tolerance. *Transportation Research Part B* 79, 16–49.
- 44 Han, K., Sun, Y., Liu, H., Friesz, T.L., Yao, T., 2015b. A bi-level model of dynamic traffic signal control with
45 continuum approximation. *Transportation Research Part C* 55, 409–431.
- 46 Han, K., Piccoli, B., Friesz, T.L., 2015c. Continuity of the path delay operator for dynamic network loading with
47 spillback. *Transportation Research Part B*, DOI: 10.1016/j.trb.2015.09.009.
- 48 Hearn, D. W., 1982. The gap function of a convex program. *Operations Research Letters* 1, 67–71.
- 49 Hearn, D.W., Ribera, J., 1980. Bounded flow equilibrium problems by penalty methods. In *Proceedings of the 1980*
50 *IEEE International Conference on Circuits and Computers*, 162–166.
- 51 Ibarra-Rojas, O.J., Delgado, F., Giesen, R., Muñoz, J.C., 2015. Planning, operation, and control of bus transport
52 systems: a literature review. *Transportation Research Part B* 77, 38–75.
- 53 Kepaptsoglou, K., Karlaftis, M., 2009. Transit route network design problem: review. *Journal of Transportation*
54 *Engineering* 135 (8), 491–505.
- 55 Lam, W.H.K., Gao, Z.Y., Chan, K.S., Yang, H., 1999. A stochastic user equilibrium assignment model for congested
56 transit networks. *Transportation Research Part B* 33, 351–368.

- 1 Larsson, T., Patriksson, M., 1995. An augmented Lagrangean dual algorithm for link capacity side constrained traffic
2 assignment problems. *Transportation Research Part B* 29, 433-455.
- 3 Larsson, T., Patriksson, M., 1998. Side constrained traffic equilibrium models—traffic management through link tolls.
4 *Equilibrium and advanced transportation modelling*. Springer US, 125-151.
- 5 LeBlanc, L.J., 1975. An algorithm for the discrete network design problem. *Transportation Science* 9 (3), 183-199.
- 6 Li, P., Mirchandani, P., Zhou, X., 2015. Solving simultaneous route guidance and traffic signal
7 optimization problem using space-phase-time hypernetwork. *Transportation Research Part B* 81(1), 103-130.
- 8 Lo, Andrew W. 2013. The origin of bounded rationality and intelligence. *Proceedings of the American Philosophical*
9 *Society* 157, 269-280.
- 10 Lou, Y., Yin, Y., Lawphongpanich, S., 2009. Robust approach to discrete network designs with demand uncertainty.
11 *Transportation Research Record* 2090, 86-94.
- 12 Lou, Y., Yin, Y., Lawphongpanich, S., 2010. Robust congestion pricing under boundedly rational user equilibrium.
13 *Transportation Research Part B* 44 (1), 15-28.
- 14 Luatkep, P., Sumalee, A., Lam, W.H., Li, Z.C., Lo, H.K., 2011. Global optimization method for mixed transportation
15 network design problem: a mixed-integer linear programming approach. *Transportation Research Part B* 45 (5),
16 808-827.
- 17 Magnanti, T.L., Wong, R.T., 1981. Accelerating benders decomposition: algorithmic enhancement and model
18 selection criteria. *Operations Research* 29 (3), 464-484.
- 19 Mahmassani, H., Chang, G., 1987. On boundedly rational user equilibrium in transportation systems. *Transportation*
20 *Science* 21 (2), 89-99.
- 21 Mahmoudi, M., Zhou, X., 2016. Finding Optimal Solutions for Vehicle Routing Problem with Pickup and Delivery
22 Services with Time Windows: A Dynamic Programming Approach Based on State-space-time Network
23 Representations. *Transportation Research Part B* 89, 19-42.
- 24 Marcotte, P., Nguyen, S., Schoeb, A., 2004. A strategic flow model of traffic assignment in static capacitated networks.
25 *Operations Research* 52 (2), 191-212.
- 26 Martello, S., Toth, P., 1990. *Knapsack Problems: Algorithms and Computer Implementation*, John Wiley and Sons.
- 27 Mokhtarian, P. L., Chen, C., 2004. TTB or not TTB, that is the question: a review and analysis of the empirical
28 literature on travel time (and money) budgets. *Transportation Research Part A* 38(9), 643-675.
- 29 Morikawa, T., Miwa, T., Kurauchi, S., Yamamoto, T., Kobayashi, K., 2005. Driver's route choice behavior and its
30 implications on network simulation and traffic assignment. *Simulation Approaches in Transportation Analysis*,
31 341-369.
- 32 Nguyen S., Pallottino, S., 1988. Equilibrium Traffic Assignment for Large Scale Transit Networks. *European Journal*
33 *of Operations Research* 37(2), 176-186.
- 34 Nguyen, S., Pallottino, S., Malucelli, F., 2001. A modeling framework for passenger assignment on a transport
35 network with timetables. *Transportation Science* 35, 238-249.
- 36 Nie, Y. M., Li, Q., Chen, P. W., 2013. Reliable Routing in Transit Networks. Technical report, Center for the
37 Commercialization of Innovative Transportation Technology, Northwestern University, Evanston, IL.
- 38 Nie, Y., Wu, X., Zissman, J., Lee, C., Haynes, M., 2010. Providing reliable route guidance: Phase ii. Technical report,
39 Center for the Commercialization of Innovative Transportation Technology, Northwestern University, Evanston,
40 IL.
- 41 Nie, Y., Zhang, M., Lee, D.H., 2004. Models and algorithm for the traffic assignment problem with link capacity
42 constraints. *Transportation Research Part B* 38 (4), 285-312.
- 43 Niu, H., Zhou, X., 2013. Optimizing urban rail timetable under time-dependent demand and oversaturated conditions.
44 *Transportation Research Part C* 36, 212-230.
- 45 Niu, H., Zhou, X., Gao, R., 2015. Train scheduling for minimizing passenger waiting time with time-dependent
46 demand and skip-stop patterns: nonlinear integer programming models with linear constraints. *Transportation*
47 *Research Part B* 76, 117-135.
- 48 Nuzzolo, A., Crisalli, U., Rosati, L., 2012. A schedule-based assignment model with explicit capacity constraints for
49 congested transit networks. *Transportation Research Part C* 20(1), 16-33.
- 50 Nuzzolo, A., Russo, F., Crisalli, U., 2001. A doubly dynamic schedule-based assignment model for transit networks.
51 *Transportation Science* 35, 268-285.
- 52 Pallottino, S., Scutellà, M. G., 1998. Shortest path algorithms in transportation models: classical and innovative
53 aspects. *Equilibrium and Advanced Transportation Modelling*, (Eds.) Marcotte, P., Nguyen, S., Kluwer Academic
54 Publishers, MA, 245-281.
- 55 Poon, M.H., Wong, S.C., Tong, C.O., 2004. A dynamic schedule-based model for congested transit networks.
56 *Transportation Research Part B* 38, 343-368.

- 1 Poorzahedy, H., Rouhani, O.M., 2007. Hybrid meta-heuristic algorithms for solving network design problem.
- 2 European Journal of Operational Research 182 (2), 578–596.
- 3 Ramming, M.S., 2001. Network knowledge and route choice. Ph.D. Thesis. Massachusetts Institute of Technology.
- 4 Rosenthal, R., 2015. GAMS: A User's Guide. GAMS Development Corporation.
- 5 Ruan, J.M., Liu, B., Wei, H., Qu, Y., Zhu, N. and Zhou, X., 2016. How Many and Where to Locate Parking Lots? A
- 6 Space–time Accessibility-Maximization Modeling Framework for Special Event Traffic Management. Urban
- 7 Rail Transit (2), 59-70.
- 8 Spiess, H., Florian, M., 1989. Optimal strategies: a new assignment model for transit networks. Transportation
- 9 Research Part B 23 (2), 83–102.
- 10 Szeto, W.Y., 2003. Dynamic Traffic Assignment: Formulations, Properties, and Extensions. PhD Thesis, The Hong
- 11 Kong University of Science and Technology, H.K.
- 12 Szeto, W.Y., Jiang, Y., 2014a. Transit assignment: Approach-based formulation, extragradient method and paradox.
- 13 Transportation Research Part B 62, 51–76.
- 14 Szeto, W.Y., Jiang, Y., 2014b. Transit route and frequency design: Bi-level modeling and hybrid artificial bee colony
- 15 algorithm approach. Transportation Research Part B 67, 235-263.
- 16 Tong, C. O., Richardson, A. J., 1984. A computer model for finding the time-dependent minimum path in a transit
- 17 system with fixed schedules. Journal of Advanced Transportation, 18(2), 145-161.
- 18 Tong, L., Zhou, X., Miller, H.J., 2015. Transportation network design for maximizing space–time accessibility.
- 19 Transportation Research Part B 81(2), 555–576.
- 20 Verbas, I., Mahmassani, H., 2015. Integrated frequency allocation and user assignment in multi-modal transit
- 21 networks: methodology and application to large-scale urban systems. Transportation Research Record 2498, 37-
- 22 45.
- 23 Wang, D.Z.W., Lo, H.K., 2010. Global optimum of the linearized network design problem with equilibrium flows.
- 24 Transportation Research Part B 44 (4), 482–492.
- 25 Wang D.Z.W., Liu H, Szeto W. Y., 2015. A novel discrete network design problem formulation and its global
- 26 optimization solution algorithm. Transportation Research Part E 79, 213-230.
- 27 Wang, S., Meng, Q., Yang, H., 2013. Global optimization methods for the discrete network design problem.
- 28 Transportation Research Part B 50, 42–60.
- 29 Xiong, Y., Schneider, J.B., 1992. Transportation network design using a cumulative algorithm and neural network.
- 30 Transportation Research Record 1364, 37–44.
- 31 Xu, W., He, S., Song, R., and Sohail, S., 2012. Finding the K shortest paths in a schedule-based transit network.
- 32 Computers & Operations Research 39(8), 1812–1826.
- 33 Xu, X., Liu, J., Li, H., and Hu, J. 2014. Analysis of subway station capacity with the use of queueing theory,
- 34 Transportation Research Part C 38, 28–43
- 35 Yang, H., Bell, M.G.H., 1998. Models and algorithms for road network design: a review and some new developments.
- 36 Transport Reviews 18 (3), 257–278.
- 37 Yang, L., Zhang, Y., Li, S., Gao, Y., 2016. A two-stage stochastic optimization model for the transfer activity choice
- 38 in metro networks. Transportation Research Part B 83, 271-297.
- 39 Yin, J., Tang, T., Yang, L., Gao, Z., Ran, B., 2016. Energy-efficient metro train rescheduling with uncertain time-
- 40 variant passenger demands: An approximate dynamic programming approach. Transportation Research Part B 91,
- 41 178-210.
- 42 Yen, J., 1971. Finding the k shortest loopless paths in a network. Management Science 17, 712–716.
- 43 Zhu, S., 2011. The Roads Taken: Theory and Evidence on Route Choice in the Wake of the I-35w Mississippi River
- 44 Bridge Collapse and Reconstruction. Ph.D. thesis, University of Minnesota.
- 45 Zhou, X., Mahmassani, H. S., Zhang, K., 2008. Dynamic micro-assignment modeling approach for integrated
- 46 multimodal urban corridor management. Transportation Research Part C 16(2), 167-186.
- 47 Zhou, X., Taylor, J., 2014. DTALite: A queue-based mesoscopic traffic simulator for fast model evaluation and
- 48 calibration. Cogent Engineering 1 (1), 961345.
- 49 Ziliaskopoulos, A., Mahmassani, H. S., 1993. A time-dependent shortest path algorithm for real-time intelligent
- 50 vehicle/highway systems applications. Transportation Research Record 1408, 94-100.