Transportation Research Part B 93 (2016) 1–16

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

Gap-based transit assignment algorithm with vehicle capacity constraints: Simulation-based implementation and large-scale application

Ömer Verbas, Hani S. Mahmassani*, Michael F. Hyland

Northwestern University, Transportation Center, 600 Foster St. 3rd Floor, Evanston, IL 60208, USA

ARTICLE INFO

Article history:
Received 6 September 2015
Revised 4 July 2016
Accepted 4 July 2016
Available online 22 July 2016

Keywords:
Transit assignment
Dynamic network assignment
User equilibrium
Large-scale networks
Gap
Simulation
Multimodal transit

ABSTRACT

This paper presents a gap-based solution method for the time-dependent transit assignment problem with vehicle capacity constraints. A two-level, simulation-based methodology is proposed, which finds the least cost hyperpaths at the upper level and performs the assignment of transit travelers on the hyperpaths at the lower level. The detailed simulation of travelers and vehicles at the lower level allows modelers to capture transit network complexities such as transfers/missed connections, receiving a seat/standing and boarding/being rejected to board. This 'hard' implementation of vehicle capacity constraints at the lower level is aggregated into 'soft constraints' at the upper level for the least cost hyperpath calculation. Using a gap-based assignment procedure, user equilibrium is reached on large-scale networks in a computationally efficient manner. The algorithm is tested on the large-scale Chicago Transit Authority network. The gap-based approach outperforms the commonly used method of successive averages approach in terms of rate of convergence and quality of results. Furthermore, sensitivity analyses with respect to network parameters illustrate the robustness of the proposed two-level solution procedure.

1. Introduction

The growing role of public transit in large cities around the world requires a more advanced understanding of transit at different decision levels. In order to evaluate several performance measures and make decisions at the tactical, strategic and planning levels, planning agencies commonly rely on assignment and simulation tools. The validity and effectiveness of an assignment algorithm relies on whether and how fast it equilibrates, as well as how realistically it captures the complexities in a transit network such as transfers, vehicle capacities etc. Since the seminal work of Beckmann et al. (1956), the user equilibrium problem has been studied extensively in the literature (Boyce et al., 2005).

Wardrop's (1952) first principle states that travelers in a transportation network choose paths/routes that minimize their own individual travel times, i.e. the travelers choose the 'shortest path' (Dijkstra, 1959) between their origin and destination. Thus at equilibrium, no individual traveler can reduce her travel time by unilaterally switching routes. Wardrop's principle has also been applied to the transit assignment problem. Dial (1967) and Le Clercq (1972) were early contributors to the transit assignment problem. Their 'single best path' assumption was later replaced by a more realistic 'optimal strategy'

* Corresponding author. Fax: +1 847 491 3090.
E-mail addresses: omer@northwestern.edu (Ö. Verbas), masmah@northwestern.edu (H. S. Mahmassani), michaelhyland2013@u.northwestern.edu (M. F. Hyland).

http://dx.doi.org/10.1016/j.trb.2016.07.002
0191-2615/

approach (Spiess and Florian, 1989) which is also referred to as the transit 'hyperpath' approach in the literature (Nguyen and Pallottino, 1988, 1989). Spiess and Florian (1989) also consider the effects of congestion in their transit assignment algorithm; however, no test results are provided. De Cea and Fernández (1993) and Wu et al. (1994) explicitly include the effects of congestion via passenger loads into their assignment algorithm; however, both algorithms include a 'soft capacity restriction' whereby travelers are never explicitly prevented from boarding a vehicle, although there is a waiting cost penalty that tends to infinity as vehicles approach their carrying capacity. Cepeda et al. (2006) use a method of successive averages (MSA) to solve the transit assignment problem with congestion effects and soft capacity restrictions. Hamdouch and Lawphongpanich (2008) include a 'hard' capacity constraint into the mathematical formulation of the problem. The authors subsequently show the existence of a solution under a hard capacity constraint.

Hamdouch et al. (2011) increase the realism and applicability of the transit assignment problem by distinguishing between seated and standing travelers in their capacity constrained analytical transit assignment model. Schmöcker et al. (2011) proposed a formulation that considers seat capacity in a spatially expanded network, whereby travelers make route choice decisions based on the probability of finding a seat on different routes (i.e. the cost associated with standing in a transit vehicle is different from sitting in a transit vehicle on one's journey). However, this formulation assumes unlimited standing capacity. Cominetti and Correa (2001) include walking links in their formulation of the transit assignment problem. These links allow travelers to walk between transit stops or even forego a transit leg if it is faster for them to walk than to wait for and ride a transit vehicle.

Transit assignment techniques are used as an integral part of service design algorithms. Han and Wilson (1982) and Baaj and Mahmassani (1990) proposed rule-based, assignment techniques used in the context of service design procedures. Abdelghany and Mahmassani (2001) presented an early simulation-assignment framework for intermodal networks; however, it only considers congestion on the road network but not on transit vehicles. The model's scalability to large-scale networks is also limited. Mahmassani et al. (2007) introduced and implemented a freight transportation simulation-assignment model which analyzes the transportation of multiproduct, intermodal freight on a large-scale intermodal rail network spanning the continent of Europe (Zhang et al., 2008).

Recently, a transit assignment logic for large-scale networks interfaced with a road highway simulator was developed (Khani et al., 2012; Khani, 2013; Noh, 2013). This line of study was further extended to include different types of transit path finding algorithms (Khani et al., 2015). Moreover, the recent works by Moyo Oliveros (2014), and Wahba and Shalaby (2011, 2014) seek to integrate transit assignment with simulation.

The method of successive averages (MSA) is one common method used to reach convergence when solving the dynamic transit assignment problem using simulation methods (Sbayti et al., 2007). The main disadvantage of MSA is that it treats all the paths equally. For a given origin, destination, and departure time triplet, the relative shift of flow to the current best path from a slightly sub-optimal path is the same as the shift of flow from a severely sub-optimal path. Sbayti et al. (2007) modified the MSA approach by sorting the vehicles based on the gap between their experienced cost and the current best cost. This reformulation guarantees that vehicles with the largest gap are forced to switch paths, while the step size of the MSA is maintained. Lu et al. (2009) further improve upon the basic MSA approach; their paper includes both flow-based and vehicle-based formulations. The flow-based formulation shifts flow from inferior paths to the current best path in a proportion relative (scaled) to the cost difference between the two paths. In their vehicle-based formulation, the probability of a vehicle switching paths is proportional to the relative gap between the vehicle's experienced path and the best path.

This paper presents a two-level, simulation-based methodology, which finds the least cost hyperpaths at the upper level and performs the assignment of transit travelers on the hyperpaths at the lower level. The study adapts the gap-based solution method proposed by Lu et al. (2009) to the time-dependent transit assignment problem with vehicle capacity constraints. The following aspects of the model are noteworthy:

- The transit network is multi-modal; besides the conventional transit modes (bus and rail), it includes walking and biking options explicitly rather than only as access or egress modes.
- The different service patterns of a transit route are captured explicitly (Verbas and Mahmassani, 2013).
- The multi-dimensionality of the network due to having multiple modes and service patterns is captured by vectorizing the links for different time intervals. This prevents the network from having to be expanded spatially and temporally.
- The assignment methodology utilizes a least cost hyperpath algorithm (Verbas and Mahmassani, 2015b) that is time-dependent and frequency-based.
- The experience of the travelers, as well as the movement and loads of vehicles are captured via a multi-agent particle simulation introduced in another study (Verbas et al., 2015). This allows the modeling of transit network complexities such as transfers/missed connections, receiving a seat/standing and boarding/being rejected to board.
- The vehicle capacity constraints are captured at two different levels. There is a 'hard' capacity constraint at the lower level, where the simulation strictly prevents individuals from boarding a full vehicle. At the upper level, there is a 'soft constraint', in which the link cost has an exponentially increasing penalty component. This penalty deters the hyperpath algorithm from including a link/pattern pair as part of the hyperpath tree (Verbas and Mahmassani, 2015b).

The remainder of the paper is structured as follows. The next section provides the model formulation, where the time-dependent equilibrium conditions are formulated as a gap-minimizing mathematical program. Section 3 introduces the transit assignment and simulation platform to solve the gap-minimizing mathematical program. The simulation of vehicles and travelers, as well as the implementation of the soft and hard capacity constraints are described in Section 4.