Transportation Research Part B 81 (2015) 103–130

Contents lists available at ScienceDirect

Transportation Research Part B

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

Solving simultaneous route guidance and traffic signal optimization problem using space-phase-time hypernetwork

Pengfei Li [a], Pitu Mirchandani [a], Xuesong Zhou [b,*]

[a] School of Computing, Informatics and Decision Systems Engineering, Arizona State University, Tempe, AZ 85281, USA
[b] School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ 85287, USA

A R T I C L E   I N F O

Article history:
Received 17 February 2015
Received in revised form 28 August 2015
Accepted 29 August 2015
Available online 14 September 2015

Keywords:
Dynamic traffic assignment
Network traffic signal control
Route guidance
Lagrangian relaxation
Dynamic programming
Optimization
Network flow modeling

A B S T R A C T

This paper addresses the problem of simultaneous route guidance and traffic signal optimization problem (RGTSO) where each vehicle in a traffic network is guided on a path and the traffic signals servicing these vehicles are set to minimize their travel times. The network is modeled as a space-phase-time (SPT) hyper-network to explicitly represent the traffic signal control phases and time-dependent vehicle paths. A Lagrangian-relaxation-based optimization framework is proposed to decouple the RGTSO problem into two subproblems: the Route Guidance (RG) problem for multiple vehicles with given origins and destinations and the Traffic Signal Optimization (TSO) problem. In the RG subproblem, the route of each vehicle is provided subject to time-dependent link capacities imposed by the solution of the TSO problem, while the traffic signal timings are optimized according to the respective link travel demands aggregated from the vehicle trajectories. The dual prices of the RG subproblem indicate search directions for optimization of the traffic signal phase sequences and durations in the TSO subproblem. Both RG and TSO subproblems can be solved using a computationally efficient finite-horizon dynamic programming framework, enhanced by parallel computing techniques. Two numerical experiments demonstrated that the system optimum of the RGTSO problem can be quickly reached with relatively small duality gap for medium-size urban networks.

1. Introduction

Most traffic signal optimization (TSO) methods, whether offline or online, assume link flows and turning ratios, also referred to as turning proportions, of the flows at intersections. Implicit assumption in most of the underlying optimization models is that the flows for the signal timing planning horizon are stationary and, thus, signal timings are determined for the given flows and turning proportions. In the consideration of the larger picture of estimating vehicle flow from origins to destinations, researchers have developed Dynamic Traffic Assignment model (DTA), further discussed in literature review in Section 2, to predict individual route of each vehicle from its origin to its destination at "equilibrium" which effectively defines stationary traffic flows. This equilibrium is generally called "user equilibrium" originally defined as the first principal of Wardrop, and in some applications "system optimal" originally defined as the second principal of Wardrop (1952). Since route travel times are essential to model travelers' route choices, the underlying network model assumes link travel times and some approximate travel times through the intersections specific to straight-through, right-turning and left-turning

* Corresponding author. Tel.: +1 480 9655827.
E-mail addresses: pengfeili@asu.edu (P. Li), pitu@asu.edu (P. Mirchandani), xzhou74@asu.edu (X. Zhou).

http://dx.doi.org/10.1016/j.trb.2015.08.011
0191-2615/

routes. Although rarely done in practice, it is possible to iteratively solve a DTA model to come up with link flows and turning proportions and then solve a TSO model to come up with signal timings for those flows and proportions. Unfortunately the approximate travel times through intersections are poorly defined in DTA models since individual vehicles may reach the intersection during various signal phases and thus travel times for the same route could have large variations.

Furthermore, with the advent of advanced computing, and significant penetration of navigation systems, it should be possible to provide both route guidance and signal service by setting appropriate signal timings for vehicles being provided route guidance. The application of the model/method developed in this paper also apply to (i) connected vehicles systems which would require simultaneous route guidance and signal service, (ii) multimodal traffic signal systems where different modes (private automobiles, transit, etc.) may be provided with differentiated routes and signal services, and (iii) automated vehicle operation where such vehicles may be required to travel in specific lanes and require specific signalization.

Briefly, in the combined Route Guidance and Traffic Signal Optimization (RGTSO) model described in this paper, the phase status of each traffic signal is modeled with variables describing the movement allowed at the scheduled phase. Link travel times depend on the number of vehicles routed through each link as per traffic flow theory. Each vehicle traveling from its origin to its destination travels each link in a candidate route with travel time depending upon the resultant vehicle flow on the link; the waiting, if any, at the intersection, explicitly depends on the queues and the signal status. The space-time trajectory of each vehicle through the network will give its total travel time and the optimization objective value becomes the total travel times of all guided vehicles which can be iteratively decreased by changing guided routes and signal phase schedules. Other optimization criteria could be considered such as minimizing number of stops to improve coordination, minimizing average queue sizes, minimizing fuel consumption, etc. Decomposition of the space-phase-time (SPT) hyper-network results in two subnetworks: namely, space-time network and phase-time network; while decomposition of the overall optimization problem results in two subproblems: TSO and route guidance (RG), where dual prices from the route guidance problem indicate search directions for the TSO subproblem. Details of the model and solution methods are provided in Sections 3–5. Some illustrative computational results are provided in Section 6. Section 7 gives some concluding remarks and points to some promising future research directions.

2. Related previous work

Traffic flow models are fundamental underlying bases for both RG and TSO problems. Specifically, vehicle routing and traffic signal performance are dependent on traffic flow attributes and traffic states of the links which, in turn, are dependent on vehicle flow propagation within networks, modeled through fundamental relationships between flow and density. As an example, Daganzo (1994, 1995) developed the cell transmission model (CTM) where network links are divided into many cells and vehicles move from cell to cell. The core part of the CTM model is a discrete approximation of the continuous Lighthill-Whitham-Richards traffic flow model proposed by Lighthill and Whitham (1955) and Richards (1956). Further review of related literature below is divided into four parts: (1) TSO models; (2) dynamic traffic assignment (DTA) models; (3) relationship between DTA and TSO models, and (4) solving combined DTA and TSO problems.

2.1. Traffic signal optimization models

The models for TSO can be further divided into two categories: off-line signal optimization and on-line (real-time) methods. Among the off-line signal optimization models, TRANSYT model (or its version in North America, TRANSYT-7F) is most used; it aims to optimize traffic signals timings along corridors or within networks. Wallace et al. (1984) describe the underlying traffic models in TRANSYT as estimating the flow propagation and resultant control delays for given link flows, turning ratios and signal timings based on macroscopic nonlinear delay models. TRANSYT uses a steepest decent algorithm for optimizing cycle lengths, splits and offsets. In PASSER-II, the coordination along corridors is optimized, referred to as bandwidth maximization, and the underlying delay model is a modified nonlinear delay model proposed by Webster (1958). As described by Chang and Messer (1991), PASSER-II uses first-order optimization methods, such as gradient descent, to obtain the optimal signal timings in terms of cycle length, phase sequencing, splits, and offsets. Macroscopic nonlinear control delay models have been defined in the Highway Capacity Manual published by National Research Council (2000) and they result in the underlying control delay model in Synchro (Husch and Albeck, 2003) for setting signal timings in a network, where the optimization approach is in essence an enumeration method to seek the optimal signal timings. MAXBAND is another off-line optimization method proposed by Little et al. (1981) to maximize the bandwidth where the model was formulated as a mixed integer linear program (MILP) problem. Later, MAXBAND was extended by Gartner et al. (1991) to a so-called multi-band approach to optimize multiple bandwidths, applied on multiple road segments and recently further extended by Zhang et al. (2015) to an asymmetrical version. In other literature, the TSO problem on networks has been formulated as: (a) other MILPs, e.g., Gartner and Stamatiadis (2002) and Han et al. (2014), (b) linear programming problems, e.g., Li and Zhang (2014),(c) microscopic-simulation-based optimization: such as the genetic-algorithm-based signal optimization by Park et al. (2001) and Stevanovic et al. (2008), "retrospective approximation" by Li et al. (2010), and emission reduction signal optimization by Osorio and Nanduri (2015) (d) reinforcement-learning-based signal optimization by Zain et al. (2009), (e) subgradient-based optimization approaches by Rinaldi and Tampère (2015) and Li et al. (2015); (f) store-and forward queues of vehicles, e.g., D'Ans and Gazis (1976), and (e) space-time networks such as of Carey and Srinivasan (1994).