Transportation Research Part C 85 (2017) 684–710

Contents lists available at ScienceDirect

Transportation Research Part C

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

A bi-level distributed approach for optimizing time-dependent congestion pricing in large networks: A simulation-based case study in the Greater Toronto Area☆

Aya Aboudina[a,b,*], Baher Abdulhai[a]

[a] Department of Civil Engineering, University of Toronto, M5S 1A4, Canada
[b] Cairo University, Faculty of Engineering, 12631 Giza, Egypt

ARTICLE INFO

Keywords:
Congestion-pricing
Optimization
Distributed computing
Dynamic traffic assignment simulation
Departure time choice
Large-scale network

ABSTRACT

Congestion pricing is one of the most widely contemplated methods to manage traffic congestion by charging fees for the use of roads, more where and when it is congested, and less where and when it is not. This study presents a bi-level distributed approach for optimal time-dependent congestion pricing in large networks. The bi-level procedure involves a theoretical model of dynamic congestion pricing and a distributed optimization algorithm. The bi-level toll optimization module is integrated into a testbed of hybrid departure time choice and dynamic traffic assignment simulation models for the Greater Toronto Area (GTA). The integrated system provides a unified (location- and time-specific) congestion pricing system that determines optimal tolling and evaluates its impact on road traffic congestion and travellers' behavioural choices, including departure time and route choices. For the system's large-scale nature and the consequent computational challenges, the optimization algorithm is executed concurrently on a parallel computing cluster. The system is applied to a simulation-based case study of tolling major highways in the GTA while capturing the network-wide regional effects of tolling. The travel demand and drivers' attributes are extracted from regional household travel survey data that reflect travellers' heterogeneity. The main results indicate that: (1) optimal variable pricing reflects congestion patterns and induces departure time re-scheduling and rerouting patterns, resulting in improved average travel times and schedule delays, (2) optimal tolls intended to manage traffic demand are significantly lower than those intended to maximize toll revenues, (3) tolled routes have different sensitivities to identical toll changes, (4) the start times of longer trips are more sensitive (elastic) to variable distance-based tolling policies compared to shorter trips, (5) toll payers benefit from tolling even before toll revenues are spent, and (6) the optimal tolling policies determined offer a win–win solution in which travel times are improved while also raising funds to invest in sustainable transportation infrastructure.

1. Introduction

Congestion pricing is widely viewed among economists as one of the promising methods to tackle traffic congestion problems (Washbrook et al., 2006). Highway agencies and roadway authorities struggle with the policy-oriented and politically driven dilemma of whether or not to toll their roads; however, this should not be the question as the merits of adopting full-cost pricing were

☆ This article belongs to the Virtual Special Issue on "Sustainable Transportation".

* Corresponding author at: Department of Civil Engineering, University of Toronto, M5S 1A4, Canada.
E-mail addresses: aya.aboudina@mail.utoronto.ca (A. Aboudina), baher.abdulhai@utoronto.ca (B. Abdulhai).

http://dx.doi.org/10.1016/j.trc.2017.10.004
Received 29 September 2016; Received in revised form 2 October 2017; Accepted 3 October 2017
0968-090X/

established decades ago (Small and Verhoef, 2007).

Road pricing has a long history, with turnpikes dating back at least to the seventeenth-century in Great Britain and the eighteenth-century in the US (Small and Verhoef, 2007). Road pricing for congestion management is more recent; it is referred to as 'congestion pricing'. The purpose of congestion pricing is to precisely target and reduce congestion when and where it typically appears. This is accomplished by charging fees for the use of certain roads in order to reduce traffic demand or distribute it more evenly over time (away from the peak period) and space (away from overly congested facilities). In other words, congestion pricing involves charging drivers for the use of roads, more where and when it is congested, and less where and when it is not (Levinson, 2016). This will reduce travel – hence congestion – on congested routes and time periods, and may increase demand on uncongested routes and time periods, where there is surplus capacity. i.e., it works towards balancing the load on the network and optimizing the overall system performance; a strategy undertaken in other transport modes such as air transport, as well as most time-sensitive services (e.g., electricity and power systems, recreation industry such as cinemas and restaurants).

The earliest modern congestion pricing application is Singapore's Area License scheme, established in 1975. Since then, other applications have appeared, varying from single facilities such as bridges or toll roads to tolled express lanes as in the US, toll cordons as in Norway, and area-wide pricing as in London. A number of cities have implemented or are in the process of implementing road pricing. Highway 407 in Toronto, which was opened to traffic in 1997, is the world's first all-electronic, barrier-free toll highway, in which tolls are charged based on vehicle type, distance driven, time of day, and day of the week (Lindsey, 2008). Except for the Highway 407 ETR, tolls in Canada do not vary over time, and no area-based road pricing scheme has been implemented in Canada, which lags behind the United States and a number of countries in Europe and Asia with respect to pricing practices.

This study is motivated to develop a bi-level distributed approach for optimal congestion pricing in large networks. The first level involves determining variable queueing-delay-eliminating toll structures for congested facilities. The second level involves iterative optimization fine-tuning of the toll structures determined in the first level to achieve the best possible network performance, while considering the route and departure time shift impacts of tolling network-wide. The second level also uses a robust iterative optimization algorithm that is run concurrently (i.e., distributed) on a parallel computing cluster. The bi-level toll optimization module is integrated into a hybrid testbed of (1) a departure time choice model and (2) a DTA simulation model of the GTA (Aboudina et al. (2016)). The integrated system provides a comprehensive tool for optimal tolling strategy determination and evaluation, while explicitly capturing departure time and route choices in a large-scale dynamic traffic assignment (DTA) simulation environment. Without belittling their probable occurrence, mode-choice responses to tolling are beyond the scope of this study, particularly considering lack of transit capacity in the GTA. Further investigation of mode related issues will be considered in future work.

The system robustness and effectiveness are examined through a simulation-based case study of tolling major highways in the Greater Toronto Area (GTA) region, namely, the Gardiner Expressway (GE), the Don Valley Parkway (DVP), and the express lanes of Highway 401. The GTA region, home of 6.4 million residents, involves widespread activities, heterogeneous travel behaviour, a wide range of socioeconomic attributes of travellers, multiple routing options, as well as many satellite cities, which make it an ideal case study in which to test any large-scale traffic control policy. The outcomes of this case study demonstrate the potential impacts of the traffic management technology, implemented herein, on the transportation system. For examples, it was found that the optimal variable pricing determined induces departure time re-scheduling and network-wide rerouting resulting in improved average travel times and schedule delays. It was also found that optimal tolls intended to manage traffic demand are significantly lower than those intended to maximize toll revenues, and that toll payers benefit from tolling even before toll revenues are spent. In other words, the congestion pricing traffic management technology presented in this paper offers a win-win solution in which travel times are improved network-wide while also raising funds to invest in sustainable transportation infrastructure.

2. Literature review

Congestion pricing of roads has been studied for decades but often at a theoretical and small scale. Further, only a few larger scale systems have been implemented and subsequently studied, such as, for example, Stockholm congestion charging system which has been well studied and analyzed via different dynamic models by Engelson et al. (2013), de Palma et al. (2014), and Saifuzzaman et al. (2016). One reason for the lack of large scale models and analyses is that most academic studies have applied simple, abstract models to small or even hypothetical networks, which do not adequately capture many real-world complexities. Further, the users' individual responses to pricing (e.g. mode and departure time choices) are usually oversimplified or neglected; and if considered, the impact of drivers' attributes on their choices was often not captured due to lack of supporting data and behavioural econometric models. (See Aboudina et al. (2016) for further details and examples of such studies).

As outlined in detail in Aboudina et al. (2016), several approaches can be followed to simulate drivers' departure time along with route changes within a traffic simulation environment. For example, (Zhu et al., 2017) proposed an approach that utilizes a mixed Bayesian Network (BN) model, rather than a utility-based model, to analyze the probabilistic relationship among mode choice, departure time choice, and other related variables. The predicted probabilities of different alternatives, in the proposed BN model, come from the calculation of posterior probabilities. (Xiong et al., 2016) proposed an integrated agent-based simulation approach. On the demand side, a naïve Bayes classifier is developed to model the decision-making. On the supply side, the integration incorporated a large-scale mesoscopic dynamic traffic simulation model. The model developed was employed in studying the en-route diversion responses of agents under real-time information provision and to analyze their impacts on network performance. The testbed developed in Aboudina et al. (2016), and used here, integrates an econometric behavioural departure time choice model (that considers both trip and driver attributes) into a large-scale DTA simulation model. This approach provides a computationally *tractable* tool to estimate departure time and route choice responses to policies affecting travel times and costs, in a *large-scale* setting.