Transportation Research Part B 71 (2015) 19–31

Contents lists available at ScienceDirect

Transportation Research Part B

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

Real-time congestion pricing strategies for toll facilities

Jorge A. Laval [a,*], Hyun W. Cho [a], Juan C. Muñoz [b], Yafeng Yin [c]

[a] School of Civil and Environmental Engineering, Georgia Institute of Technology, United States
[b] Department of Transport Engineering and Logistics, Pontificia Universidad Católica de Chile, Chile
[c] Department of Civil and Coastal Engineering, University of Florida, United States

ARTICLE INFO

Article history:
Received 15 October 2013
Received in revised form 24 September 2014
Accepted 25 September 2014
Available online 12 November 2014

Keywords:
System optimum
User equilibrium
Congestion pricing

ABSTRACT

This paper analyzes the dynamic traffic assignment problem on a two-alternative network with one alternative subject to a dynamic pricing that responds to real-time arrivals in a system optimal way. Analytical expressions for the assignment, revenue and total delay in each alternative are derived as a function of the pricing strategy. It is found that minimum total system delay can be achieved with many different pricing strategies. This gives flexibility to operators to allocate congestion to either alternative according to their specific objective while maintaining the same minimum total system delay. Given a specific objective, the optimal pricing strategy can be determined by finding a single parameter value in the case of HOT lanes. Maximum revenue is achieved by keeping the toll facility at capacity with no queues for as long as possible. Guidelines for implementation are discussed.

1. Introduction

There are currently more than a dozen cities around the world that implement zone- or cordon-based congestion pricing, and around 20 toll facilities in the United States subject to congestion pricing. The pricing strategies in these facilities are inspired by the "first-best toll" concept borrowed from the economics literature, which can be stated as "System Optimum (SO) will be equivalent to User Equilibrium (UE) with tolls derived from the SO solution" (see e.g. Carey and Watling, 2012). This concept has been adapted to the case of traffic flow rather directly, and it is our view that some important traffic dynamics properties may have been overlooked in doing so.

Although there are a number of studies examining the performance of High Occupancy Toll (HOT) lanes (see e.g. Supernak et al., 2003; Supernak et al., 2002a; Supernak et al., 2002b; Burris and Stockton, 2004; Zhang et al., 2009) and travelers' willingness to pay (Li, 2001; Burris and Appiah, 2004; Podgorski and Kockelman, 2006; Zmud et al., 2007; Finkleman et al., 2011), SO tolling policies have received little attention. Existing studies focused on ad hoc objectives that the tolling agencies may seek to achieve, such as ensuring free-flow conditions on HOT lane. For example, Li and Govind (2002) developed a toll evaluation model to assess the optimal pricing strategies of the HOT lane that can accomplish different objectives such as ensuring a minimum speed on the HOT lane, or in the general-purpose lanes (GPL), or maximizing toll revenue. Zhang et al. (2008) proposed the logit model to estimate dynamic toll rates of the HOT lane after calculating the optimal flow ratios by using feedback-based algorithm on the basis of keeping the HOT lane speed higher than 45 mph. Yin and Lou (2009) explored two approaches including feedback and self-learning methods to determine dynamic pricing strategies for the

* Corresponding author. Tel.: +1 (404) 894 2360; fax: +1 (404) 894 2278.
E-mail address: jorge.laval@ce.gatech.edu (J.A. Laval).

http://dx.doi.org/10.1016/j.trb.2014.09.015
0191-2615/© 2014 Elsevier Ltd. All rights reserved.

HOT lane, and the comparative results showed that the self-learning controller is superior to the feedback controller in view of maintaining a free-flow traffic condition for managed lanes. Lou et al. (2011) further developed the self-learning approach in Yin and Lou (2009) to incorporate the effects of lane changing using the hybrid traffic flow model in Laval and Daganzo (2006). Burris et al. (2009) examined the potential impacts of different tolling strategies on carpools, which includes removing or reducing the preferential treatment for them in the HOV lane.

In our formulation the social cost to be minimized is total system delay, and does not include the effects that tolls may have on trip generation or departure-time choice. The proposed pricing strategies are real-time, in the sense that they respond to real-time traffic arrivals in a way that minimizes total system delay for that particular rush hour. Therefore, the underlying assumption is that demand is inelastic within the day, but it could very well be elastic from day to day. In this context, this paper proposes a real-time pricing mechanism that is consistent with known properties of marginal costs under inelastic demands, i.e.: the cost of adding an additional user to a specific alternative is given by the time until congestion clears, it is not well defined when capacity is reached, and the SO assignment is not unique (Muñoz and Laval, 2005; Kuwahara, 2007). Towards this end, this paper is organized as follows. Section 2 presents the problem formulation along with the SO and UE solutions. Section 3 summarizes the general properties of SO tolls, including expressions for delays and revenue. Section 4 examines the special case of HOT lanes, and finally Section 5 presents a discussion.

2. Problem formulation

Consider the equilibrium between two alternatives with finite capacity, one of which is priced. To fix ideas, we take the example of a Managed Lane (ML) competing with the general-purpose lanes (GPL), but the formulation to be developed also applies to other cases such as toll roads or zone-based pricing. Our focus is on real-time pricing strategies and therefore we do not assume that traffic demand is known in advance, but only as it realizes.

Let $A(t)$ be the cumulative number of vehicles at time t that have entered a freeway segment containing a ML entrance, and let the corresponding flow be $\lambda(t) = \dot{A}(t)$. All vehicles are bound for a single destination past a GPL bottleneck of capacity μ_0, which may be bypassed by paying a toll $\pi(t)$ upon entering the ML at time t, which has a bottleneck of capacity μ_1; see Fig. 1.

The cumulative count curve of vehicles using route r ($r = 0$ for the GPL and $r = 1$ for the ML) is denoted $A_r(t)$ and the flow, $\lambda_r(t) = \dot{A}_r(t)$. Clearly,

$$\lambda(t) = \lambda_0(t) + \lambda_1(t), \tag{1}$$

and is assumed unimodal. Let $\tau_r(t)$ be the trip time in route r experienced by a user arriving at time t:

$$\tau_r(t) = \tau_r + w_r(t), \tag{2}$$

where τ_r is the free-flow travel time, and $w_r(t)$ is the queuing delay, which can be expressed as:

$$w_r(t) = \frac{A_r(t) - A_r(t_r)}{\mu_r} - (t - t_r), t_r < t < T_r, \tag{3}$$

where t_r and T_r represent the times when route r begins and ends being congested, respectively. Let:

$$\Delta = \tau_0 - \tau_1 \tag{4}$$

be the extra free-flow travel time for using the free alternative. Although in many cases one would expect $\tau_0 \approx \tau_1$, this will not be assumed here for maximum generality. To simplify the exposition, we assume that $\Delta > 0$ hereafter; the other two cases will be discussed in the last section of this paper. Under this assumption, we will see that $t_1 < t_0$ in the SO solution, i.e. the ML is used at capacity before the GPL, as shown next.

2.1. System optimum

The SO solution to our problem (without pricing) is presented in Fig. 2, which shows the system input–output diagram using total arrivals $A(t) = A_0(t) + A_1(t)$ and total virtual departures $D^*(t)$. [1] The area between these curves is the total system delay, i.e. the total time spent queuing in the system. The method to obtain the curve $D^*(t)$ was introduced in Muñoz and Laval (2005), and is best visualized by imagining a ring connected to the rightmost end of $D^*(t)$ that is slid along $A(t)$ from right to left until $D^*(t)$ "touches" $A(t)$ again (at point "1" in the figure). This point corresponds to the time when both alternatives start being used at capacity (t_0 in our case since $\Delta > 0$, and $\lambda(t_0) = \mu_0 + \mu_1$), and from here one can identify the arrival time of the last vehicles to experience delay in each alternative, $T_r, r = 0, 1$, and the time when the shorter alternative starts being used at capacity, (t_1 in our case, and $\lambda(t_1) = \mu_1$); see Fig. 2. This figure also shows how to obtain the total system departure curve $D(t)$, which gives the count of vehicles reaching the destination at time t. Notice that total arrivals and departures in the system are not first-in-first-out. The resulting flow pattern is summarized below (Muñoz and Laval, 2005):

System Optimum Conditions: The SO assignment when $\Delta > 0$ for users arriving at t satisfy:

1. $0 \leqslant t \leqslant t_1$: everybody uses the ML.

[1] Virtual departures are defined as the arrival curve shifted to the right by the free-flow travel time.