Dynamic Pricing in Discrete Time Stochastic Day-to-Day Route Choice Models

Tarun Rambha and Stephen D. Boyles

Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin

Abstract

The traffic assignment problem is primarily concerned with the study of user equilibrium and system optimum and it is often assumed that travelers are perfectly rational and have a complete knowledge of network conditions. However, from an empirical standpoint, when a large number of selfish users travel in a network, the chances of reaching an equilibrium are slim. User behavior in such settings can be modeled using probabilistic route choice models which define when and how travelers switch paths. This approach results in stochastic processes with steady state distributions containing multiple states in their support. In this paper, we propose an average cost Markov decision process model to reduce the expected total system travel time of the logit route choice model using dynamic pricing. Existing dynamic pricing methods in day-to-day network models are formulated in continuous time. However, the solutions from these methods cannot be used to set tolls on different days in the network. We hence study dynamic tolling in a discrete time setting in which the system manager collects tolls based on the state of the system on previous day(s). In order to make this framework practical, approximation schemes for handling a large number of users are developed. A simple example to illustrate the application of the exact and approximate methods is also presented.

Keywords: day-to-day dynamics; dynamic tolls; average cost MDP; state space aggregation

1 Introduction

Urban transportation planning is traditionally carried out using a four-step process. The first three steps are used to estimate the number of travelers/users, their origin-destination (OD) pairs, and their mode of travel. The final step, also called *route choice* or *traffic assignment*, involves assigning travelers to different routes. This assignment procedure is done assuming that traffic networks are in a state of *user equilibrium* (UE) or *Nash equilibrium* (NE) due to selfish choices made by travelers (Wardrop, 1952; Nash, 1951). Many efficient algorithms exist for finding the UE solution to the traffic assignment problem (TAP) (Larsson and Patriksson, 1992; Jayakrishnan et al., 1994; Bar-Gera, 2002; Dial, 2006; Mitradjieva and Lindberg, 2013). Another state typically of interest is called the *system optimum* (SO) in which the sum total of travel time experienced by all travelers, also called the *total system travel time* (TSTT), is minimized.

In UE models, it is often assumed that travelers are rational and have a perfect knowledge of the network topology and its response to congestion. However, when a large number of travelers interact, the extent of reasoning required to arrive at an equilibrium solution is beyond one's human ability. Two alternate concepts which do not rely on these assumptions exist in literature – *stochastic user equilibrium* (SUE) and *day-to-day dynamic models or Markovian traffic assignment models*. Both these approaches infuse uncertainty into travelers' choices and the uncertainty is assumed to result from randomness in users'

perceived travel times. However, they differ from each other in a vital way. Stochastic user equilibrium models (Dial, 1971; Daganzo and Sheffi, 1977; Sheffi, 1985), which are formulated as fixed point problems, define equilibrium as a state in which users' perceived travel times are minimized.

On the other hand, day-to-day models (Cascetta, 1989; Friesz et al., 1994; Cantarella and Cascetta, 1995) are deterministic or stochastic dynamic processes in which states/feasible flows evolve over time. In discrete time models with stochastic dynamics, travelers select paths each day based on historical information of network travel times and a probabilistic route choice mechanism which induces transitions from one state to another. Under certain mild conditions, the stochastic process can be modeled as a Markov chain with a unique steady state distribution. Thus, although the system is never at rest, it attains an 'equilibrium' in the probability distribution of flow patterns.

Since paths are selected randomly on each day, the total system travel time is no longer deterministic but is a random variable. Using the steady state distribution of the stochastic process, one can compute the expected TSTT, which can be used as a metric for studying the extent of congestion in the network. An immediate question of interest is the following. Just as congestion pricing is used to achieve SO flows in traffic assignment, can a system manager do the same to reduce the expected TSTT? Traditional congestion pricing is based upon marginal prices (Pigou, 1920). Congestion pricing helps reduce the TSTT as the UE flow on the network with marginal tolls results in a SO flow pattern in the original network. However, in a day-to-day setting, marginal prices are of little relevance. In fact, in some cases, they can result in increased TSTT as illustrated by the following example.

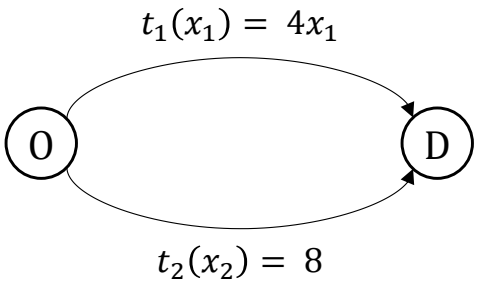


Figure 1: Sub-optimality of marginal prices in a day-to-day setting

Consider two travelers from O to D in the network shown in Figure 1. Let the vector (x_1, x_2) denote the state of the system, where x_1 and x_2 denotes the number of travelers on the top and the bottom path respectively. The above system has three states $(2, 0)$, $(0, 2)$, and $(1, 1)$, which we will refer to as states 1, 2 and 3 respectively. It is easy to verify that state 1 is a NE and state 3 is SO. Suppose both travelers use the logit choice model to select paths on each day, in which the probability of choosing the top and bottom paths are $\frac{\exp(-t_1(x_1))}{\exp(-t_1(x_1))+\exp(-t_2(x_2))}$ and $\frac{\exp(-t_2(x_2))}{\exp(-t_2(x_2))+\exp(-t_2(x_2))}$, where $t_1(x_1)$ and $t_2(x_2)$ represents the travel times as a function of the previous day's flow. The stochastic process is Markovian and the steady state probabilities of observing states 1, 2, and 3 are 0.5654, 0.1414, and 0.2932 respectively. Thus the expected TSTT is 16(0.5654) + 16(0.1414) + 12(0.2932) = 14.8272. Now suppose we price the network using marginal tolls (4 units on the top link and no toll on the bottom one) and assume that both travelers now

replace the travel time functions in the logit choice model with generalized costs (travel time + toll). The steady state distribution of the Markov chain for states 1, 2, and 3 is 0.467, 0.467, and 0.066 respectively and the expected TSTT is 15.736 which is higher than before.

Selecting the right tolls in a day-to-day setting would thus require us to estimate the steady state distribution for each admissible toll pattern and select one that minimizes the expected TSTT. However, one can do better than such static tolling schemes by dynamically varying tolls. While dynamic tolling has received some attention in literature, most existing research focuses primarily on continuous time models. These studies use control theory to determine the optimal time varying toll as the system state evolves with time according to some deterministic dynamic (Friesz et al., 2004; Xiao et al., 2014). However, continuous time formulations are not really 'day-to-day' models and their solutions cannot be used to dynamically price a network over different days. A major contribution of this paper is in addressing this gap by proposing a dynamic day-to-day pricing mechanism in a discrete time setting that computes the optimal link tolls to reduce the expected TSTT. We formulate this problem as an infinite horizon average cost Markov decision process (MDP) and seek stationary policies that inform us the tolls as a function of the state of the system. In other words, the system manager observes the state or flow pattern and sets tolls, which are then revealed to the travelers. Travelers pick paths the next day in a probabilistic manner depending on the current state and revealed tolls.

Tolls in real world transportation networks are largely levied on freeways and hence the path choice set for travelers may be assumed to be small. However, even in sparse networks, presence of a large number of travelers results in an exponential number of states. Therefore, as with most MDPs, we are faced with the *curse of dimensionality* that prevents us from using this model on practical networks. To address this, we also propose simple approximation techniques using state space aggregation to handle instances with large number of travelers and demonstrate its performance on a small test network. For most part, we will restrict our attention to a single OD pair and the logit choice model for route selection. Extensions to more general settings are conceptually straightforward.

The rest of this paper is organized as follows. In Section 2, we describe the two approaches (discrete and continuous) to model the evolution of traffic as a stochastic process. We also discuss existing literature on dynamic pricing. Section 3 describes an average cost MDP model for finding a dynamic pricing policy that minimizes the expected TSTT. In Section 4, we propose an approximate dynamic programming method using state space aggregation and test its performance on a simple network in Section 5. In Section 6 we formulate other variant MDPs such as those that optimize the probability of convergence to a flow solution and those that involve incentives. Finally, we summarize our findings and discuss pointers for future research on this topic.

2 Literature review

Day-to-day traffic models can be classified into two categories – *discrete* and *continuous*. Both these categories of models appear in literature in two flavors – *deterministic* and *stochastic*. The nomenclature is sometimes misleading as continuous time route switching processes are also referred to as 'day-to-day'

models. In this section we review existing literature on these models and dynamic pricing. The reader may refer to Watling and Hazelton (2003) and Watling and Cantarella (2013) for a more comprehensive summary of day-to-day dynamics and for their connections with UE and SUE.

2.1 Discrete time day-to-day models

Cascetta (1989) formulated the evolution of traffic as a discrete time Markov chain. The number of travelers was assumed to be finite. Extensions to model an infinite number of travelers also exist (see Watling and Cantarella, 2013). Under the following assumptions on the path choice probabilities, it was shown that the stochastic process has a unique steady state distribution: (1) time invariant path choice probabilities, (2) the probability of selecting any path between an OD pair is non-negative, and (3) the probability of choosing a path depends on the states (flow patterns) of the system on at most m previous days (which ensures that the process is m-dependent Markovian). Commonly used path choice models in literature include the logit and probit choice models. In logit choice models, the probability of selecting a path is additionally assumed to depend on a parameter θ which defines the extent of making a mistake or the extent of irrationality.

This model was extended by Cascetta and Cantarella (1991) to account for within day fluctuations in traffic. Travelers were assumed to have access to travel time information in periods prior to their departure and condition their choices based on historic day-to-day information and also within-day information. Watling (1996) studied day-to-day models for asymmetric traffic assignment problems (i.e., for ones in which the Jacobian of the cost functions is not symmetric and multiple equilibria may exist). The stationary distributions in such problems were found to have multiple modes at the stable equilibria or a unimodal shape if one of the equilibrium dominates the others.

Several efforts have been made to estimate the expected route flows and the correlations among flow patterns in day-to-day models (Davis and Nihan, 1993; Hazelton and Watling, 2004) as the computation of steady state distributions of Markov chains for cases with a large number of travelers can be intensive even when using Monte Carlo simulations. For networks with a large number of travelers, the expected flows may be approximated to an SUE solution (Davis and Nihan, 1993). Discrete time day-to-day models (see Cantarella and Cascetta, 1995) with deterministic dynamics have also been studied in literature. These models employ a deterministic mapping, e.g., best response mechanism (see Brown, 1951; Robinson, 1951), that provides the state of the system on the next day as a function of the flows observed on previous days.

2.2 Continuous time day-to-day models

Continuous time day-to-day dynamics may also be modeled as continuous time Markov chains in a manner similar to discrete time day-to-day models. This approach is relatively more common in behavioral economics (see Sandholm, 2010, Chapters 10-12). Travelers are assumed to be atomic (i.e., flows are integral) and their choices are characterized by − inertia, myopic behavior, and mutations. Inertia implies that travelers do not frequently change paths but do so only when they are given a strategy revision opportunity, which presents itself at random times. The sojourn times for each traveler (time between two successive revision opportunities) are assumed to be exponentially distributed. Myopic behavior implies that travelers

choose actions to optimize their present travel times rather than infinite-horizon discounted travel times. Mutations reflect the assumption that travelers may "tremble" or make mistakes while choosing a path. Depending on the probabilities that are assigned to the strategies that are not best responses, different learning algorithms can be constructed (Young, 1993, 2004; Kandori et al., 1993; Kandori and Rob, 1995; Blume, 1996). Under certain assumptions, existence of a unique steady state/limiting distribution can be ensured. Blume (1996) showed that for logit choice models, as θ tends to zero, the set of states with positive limiting probabilities (called a *stochastically stable set*) coincides with the set of NE. Further, for cases with a large number of players, it may be shown that deterministic and stochastic approaches are equivalent to each other when observed for a finite period of time (Sandholm, 2010).

The deterministic version of continuous day-to-day models assumes that the state of the system evolves with time as an ordinary differential equation and has been widely studied in transportation literature. Travelers are usually assumed to be infinitely divisible (non-atomic). One of the most commonly used dynamic is the Smith's dynamic (Smith, 1979) in which users shift between routes at a rate proportional to the difference between their current travel times. Other deterministic dynamics that have appeared in literature in transportation and behavioral economics include replicator dynamics (Taylor and Jonker, 1978; Smith and Price, 1973), projection dynamics (Nagurney and Zhang, 1997), and Brown-von Neumann-Nash dynamic (Brown and Von Neumann, 1950). The common objective in studying these models is to verify if the rest points of the dynamic are unique and coincide with the UE solution. Most deterministic dynamical systems are formulated using path flows. However, from a practical standpoint, as the number of paths may increase exponentially with the network size, researchers have recently developed more practical link based dynamic models (Zhang et al., 2001; He et al., 2010; Han and Du, 2012; Guo et al., 2015). In this context, Yang and Zhang (2009)and Guo et al. (2013) proposed a class of dynamic route choice models called rational adjustment processes (whose stationary states are at UE) in continuous and discrete time settings respectively.

2.3 Dynamic pricing

A considerable amount of literature on dynamic tolling exists. In the context of day-to-day models, existing methods usually focus on continuous time versions and are formulated as optimal control problems (Wie and Tobin, 1998; Friesz et al., 2004; Yang, 2008). The system is assumed to be tolled for a finite period of time and, using boundary conditions, it is ensured that the network remains in a SO state at the end of the finite time horizon. Friesz et al. (2004) developed such a pricing model in which the objective was to maximize net social benefits while ensuring that a minimum revenue target is met. Several other alternate objectives may be modeled using this framework such as minimizing travel costs and minimizing the time required to guide the system to an SO state (Xiao et al., 2014). Other related pricing literature includes studies to achieve SO flows using static tolls when users route according to SUE (Smith et al., 1995; Yang, 1999); piecewise constant pricing mechanisms that ensure the convergence of multiplicative update rule and replicator dynamics to an SO state (Farokhi and Johansson, 2015); self-learning and feedback learning controller for tolling managed lanes (Yin and Lou, 2009); and time-varying tolls in dynamic traffic assignment or within-day dynamic models (Joksimovic et al., 2005).

2.4 Summary

As noted in the previous subsections, there is a huge body of literature on different versions of day-to-day dynamic models. The congestion pricing methods developed in this paper optimize network performance in the presence of daily fluctuation in traffic flows. Hence, we use a discrete time stochastic day-to-day dynamic model along the lines of those developed by Cascetta (1989).

3 Dynamic pricing – Average cost MDP formulation

In this section, we introduce the four components of the MDP − the state space, the action space, transition probabilities and the costs and discuss a commonly used method for solving it.

3.1 Preliminaries

We make the following assumptions for the day-to-day traffic model with tolls:

1. The network has a single origin-destination (OD) pair with r routes. The formulation may be extended to include multiple OD pairs but has been avoided to simplify the notation.
2. There are a total of n travelers (assumed atomic and finite). Throughout this paper, the number of travelers will be assumed to be fixed. Although finiteness is limiting because of demand uncertainty in networks, the treatment of models with elastic demand is a topic in itself and can be justly studied only if the problem with a fixed number of travelers is fully understood.
3. Tolls can be collected in discrete amounts and along routes in the network. The discretization is mainly to assist the computation of optimal tolling policies and is realistic since currencies have a lowest denomination. The assumption that tolls are collected at a route level is not restrictive because the problem may be easily reformulated using link level tolls (in which case the action space is a vector of link tolls).
4. Users make route choice decisions on a day-to-day basis based on a given route choice model in which travel times are replaced by generalized costs. Decisions are conditioned on the previous day's travel times. All travelers are also assumed to be homogenous with the same value of travel time.
5. The objective of the system manager is to minimize the expected TSTT. Other objectives that may be of interest are discussed in Section 6.1.

State space

Let $R = \{1, \ldots, r\}$ denote the set of routes. We define the state space S as the set of all feasible route flow solutions, i.e., $\left\{(x_1, x_2, \ldots, x_r) \in \mathbb{Z}_+^r : \sum_{i \in R} x_i = n\right\}$. The vector $\mathbf{x} = (x_1, x_2, \ldots, x_r) \in S$ contains the flows on each of the paths between the OD pair. Since we are dealing with a network with r routes and n travelers, there are a total of $\binom{n+r-1}{n}$ feasible flow solutions/states. We use $\mathbf{x}_k$ to denote the state of the system at time step/day k.

Action space

We will represent an action using a toll vector $\mathbf{u} = (u_1, u_2, \ldots, u_r)$, which denotes the tolls on the paths in the network. Assume that the action space at state $\mathbf{x}$ is $U(\mathbf{x})$. Also suppose that the action space for each state $\mathbf{x}$ is the Cartesian product $\{\tau_1, \tau_2, \ldots, \tau_l\}^r$ where $\tau_1, \ldots, \tau_l$ are some allowable prices.

Transition probabilities

Let $t_i : S \to \mathbb{R}$ be the travel time on path $i \in R$ as a function of the state. We assume that the travel time functions are bounded. No further assumptions such as separability or monotonicity are needed. We suppose that the path choice probability $q_r(\mathbf{x}, \mathbf{u})$ for each traveler is a function of $\mathbf{x}$ and $\mathbf{u}$ and is positive for all routes for all state-action pairs. We further suppose that each traveler independently chooses a path using this distribution. Thus, the probability of moving from state $\mathbf{x}$ to $\mathbf{y}$ when action $\mathbf{u} \in U(\mathbf{x})$ is taken in state $\mathbf{x}$ is given by the following multinomial probability distribution

$$\Pr\left[\mathbf{y} = (y_1, y_2, \ldots, y_r)|\mathbf{x}, \mathbf{u}\right] = p_{\mathbf{xy}}(\mathbf{u}) = \frac{n!}{y_1!y_2!\ldots y_r!} q_1(\mathbf{x}, \mathbf{u})^{y_1} \ldots q_r(\mathbf{x}, \mathbf{u})^{y_r} \tag{1}$$

If travelers use the logit choice model with parameter θ, we may write $q_r(\mathbf{x}, \mathbf{u})$ as follows:

$$q_r(\mathbf{x}, \mathbf{u}) = \frac{e^{-\theta[t_r(\mathbf{x})+u_r]}}{\sum_{i=1}^{r} e^{-\theta[t_i(\mathbf{x})+u_i]}} \tag{2}$$

where $t_i(\mathbf{x}) + u_i$ is the generalized cost on route i. Then, the transition probabilities take the form

$$p_{\mathbf{xy}}(\mathbf{u}) = \frac{n!}{y_1!y_2!\ldots y_r!} \left(\frac{e^{-\theta[t_1(\mathbf{x})+u_1]}}{\sum_{i=1}^{r} e^{-\theta[t_i(\mathbf{x})+u_i]}}\right)^{y_1} \cdots \left(\frac{e^{-\theta[t_r(\mathbf{x})+u_r]}}{\sum_{i=1}^{r} e^{-\theta[t_i(\mathbf{x})+u_i]}}\right)^{y_r} \tag{3}$$

$$= \frac{n!}{y_1!y_2!\ldots y_r!} \prod_{j=1}^{r} \left(\frac{e^{-\theta[t_j(\mathbf{x})+u_j]}}{\sum_{i=1}^{r} e^{-\theta[t_i(\mathbf{x})+u_i]}}\right)^{y_j} \tag{4}$$

Remark. *Route choice processes in day-to-day models can be made more general than what has been described above. The system state on day k usually includes historical information and is defined as a vector of flows on previous m days* $\left(\mathbf{x}_k, \mathbf{x}_{k-1}, \ldots, \mathbf{x}_{k-(m-1)}\right)$*. Travelers are assumed to compute perceived travel times* $\tilde{t}_i(.)$ *for each path i between their OD pair as the sum of a weighted average of travel times on route i on previous m days and a random term that accounts for perception errors or unobserved factors.*

$$\tilde{t}_i\left(\left(\mathbf{x}_k, \ldots, \mathbf{x}_{k-(m-1)}\right)\right) = \sum_{j=0}^{m-1} w_j t_i\left(\mathbf{x}_{k-j}\right) + \tilde{\epsilon} \tag{5}$$

The terms w_j *represent the weights associated with the observed travel times on previous days. The probability of choosing path i is thus given by*

$$\Pr\left[\tilde{t}_i\left(\left(\mathbf{x}_k, \ldots, \mathbf{x}_{k-(m-1)}\right)\right) < \tilde{t}_{i'}\left(\left(\mathbf{x}_k, \ldots, \mathbf{x}_{k-(m-1)}\right)\right) \; \forall\, i \neq i', i' \in R\right] \tag{6}$$

Depending on the assumed distributions of the error terms, different route choice models such as logit and probit may be obtained. Logit choice models are relatively widely used as the path choice probabilities have a closed form expression.

Costs

Let $g(\mathbf{x}, \mathbf{u})$ denote the expected cost incurred every time decision $\mathbf{u}$ is taken in state $\mathbf{x}$. In order to minimize the expected TSTT, we define the cost as $g(\mathbf{x}, \mathbf{u}) = \sum_{\mathbf{y} \in S} p_{\mathbf{xy}}(\mathbf{u})\text{TSTT}(\mathbf{y})$, where $\text{TSTT}(\mathbf{y})$ is the total system travel time of state $\mathbf{y}$. Note that $g(\mathbf{x}, \mathbf{u})$ is bounded because the travel time functions t_i are assumed

to be bounded.

3.2 Objective and algorithms

The system manager observes the state on a particular day and chooses the tolls based on some policy $\left(\boldsymbol{\mu}(\mathbf{x})\right)_{\mathbf{x}\in S}$, which specifies the action $\boldsymbol{\mu}(\mathbf{x}) \in U(\mathbf{x})$ to be taken when in state $\mathbf{x}$ and reveals them to the travelers before the next day. Travelers make decisions based on the previous day's state and the revealed tolls as shown in Figure 2.

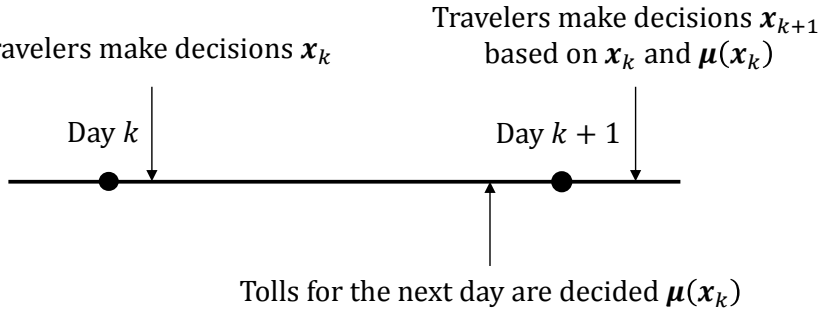


Figure 2: Timeline for the pricing mechanism

Decisions are made over an infinite horizon but at discrete intervals of time $0, 1, 2, \ldots, k, \ldots$. Let $J_{\boldsymbol{\mu}}(\mathbf{x})$ be the average cost per stage or the expected TSTT for policy $\boldsymbol{\mu}$ assuming that the system starts at state $\mathbf{x}$, i.e., $\mathbf{x}_0 = \mathbf{x}$. Thus, we may write

$$J_{\boldsymbol{\mu}}(\mathbf{x}) = \lim_{K\to\infty} \frac{1}{K}\mathbb{E}\left\{ \sum_{k=0}^{K-1} g\left(\mathbf{x}_k, \boldsymbol{\mu}(\mathbf{x}_k)\right) \,\middle|\, \mathbf{x}_0 = \mathbf{x} \right\} \tag{7}$$

Remark. *A majority of infinite horizon MDPs are formulated as discounted cost problems. In this paper, assuming that the system manager minimizes the expected TSTT, we use the average cost version instead for a couple of reasons. First, average cost MDPs are mathematically attractive as their objective values, as we will see shortly, do not depend on the initial state of the system. On the other hand, discounted cost problems are often extremely sensitive to both initial conditions and the discount factor.*

Second, and more importantly, discounted cost models are appropriate when the costs associated with state-action pairs have an economic value. In such cases, the discount factor can simply be set to the interest rate. Although estimating the monetary value of system wide travel time savings in transportation networks appears difficult, one may still use the discounted cost framework to place more weight on near-term savings in TSTT, especially when the Markov chains associated with optimal average cost policies visits states with high TSTT initially and converges to states with low TSTT only after a long time. However, for the problem instances we tested (see Section 5), the Markov chains associated with the optimal policies were found to mix quickly and the time averages of TSTT over a finite number of initial days for different sample paths were fairly close to the optimal expected TSTT, and thereby did not motivate the need for discounting.

We restrict our attention to time-invariant or stationary policies (since we are only dealing with stationary policies, the above limit always exists). The advantages of stationary policies are two-fold. First, an optimal stationary policy is relatively easy to compute. Second, since the policies do not directly depend on the day k, implementing a stationary policy is much easier. Note that stationarity of policies does not imply that the tolls are static. It implies that the tolls are purely a function of the state and as the states of the network vary over time, so do the tolls. We seek an optimal policy $\boldsymbol{\mu}^*$ such that

$$J^*(\mathbf{x}) \equiv J_{\boldsymbol{\mu}^*}(\mathbf{x}) = \min_{\boldsymbol{\mu} \in \Pi} J_{\boldsymbol{\mu}}(\mathbf{x}) \tag{8}$$

where Π denotes the set of all admissible policies. We now state some standard results concerning average cost per stage MDPs that are relevant to the current paper. A more detailed account of these can be found in Puterman (2005) and Bertsekas (2007).

Definition 1. *For a given stationary policy $\boldsymbol{\mu}$, state $\mathbf{y}$ is said to be accessible from $\mathbf{x}$, and is denoted by $\mathbf{x} \to \mathbf{y}$, if for some $k > 0$, $\mathbf{y}$ can be reached from $\mathbf{x}$ with positive probability in k days, i.e., $\Pr[\mathbf{x}_k = \mathbf{y} | \mathbf{x}_0 = \mathbf{x}, \boldsymbol{\mu}] > 0$. Further, if $\mathbf{x} \to \mathbf{y}$ and $\mathbf{y} \to \mathbf{x}$, we say that $\mathbf{x}$ communicates with $\mathbf{y}$. If $\mathbf{y}$ is not accessible from $\mathbf{x}$, we denote it by $\mathbf{x} \nrightarrow \mathbf{y}$.*

Definition 2. *For a given stationary policy $\boldsymbol{\mu}$, a subset of states $S' \subseteq S$ is a recurrent class or a closed communicating class if*

(a) All states in S' communicate with each other.

(b) $\mathbf{x} \in S'$ and $\mathbf{y} \notin S' \Rightarrow \mathbf{x} \nrightarrow \mathbf{y}$.

Definition 3. *An MDP is said to be ergodic if the Markov chain induced by every deterministic stationary policy is irreducible, i.e., has a single recurrent class.*

For the logit choice model described in this paper, the path choice probabilities and the transition probabilities between every pair of states, defined using (2) and (4) respectively, are positive for all policies. Thus, using Definitions 1 and 2, we conclude that all states communicate with each other and belong to a single recurrent class. Therefore, by Definition 3, the MDP is ergodic.

Proposition 1 (Equal costs)**.** *If an MDP is ergodic then the average cost problem has equal costs, i.e,*

$$J^*(\mathbf{x}) = J^*(\mathbf{y}), \, \forall \, \mathbf{x}, \mathbf{y} \in S \tag{9}$$

Proof. Consider a stationary policy $\boldsymbol{\mu}$. Clearly, the cost incurred up to a finite number of stages do not matter when computing the expected TSTT $J_{\boldsymbol{\mu}}(\mathbf{x})$ under the policy $\boldsymbol{\mu}$ assuming that we start at state $\mathbf{x}$, i.e., suppose $K' < \infty$, then

$$\lim_{K \to \infty} \frac{1}{K} \mathbb{E} \left\{ \sum_{k=0}^{K'-1} g\left(\mathbf{x}_k, \boldsymbol{\mu}(\mathbf{x}_k)\right) \,\middle|\, \mathbf{x}_0 = \mathbf{x} \right\} = 0 \tag{10}$$

Suppose the random variable $\widetilde{K}$ represents the time taken for the Markov chain to move from $\mathbf{x}$ to $\mathbf{y}$ for the first time under policy $\boldsymbol{\mu}$. Since the state $\mathbf{y}$ is accessible from $\mathbf{x}$ under the policy $\boldsymbol{\mu}$, it follows that

$\mathbb{E}[\widetilde{K}] < \infty$. Therefore, using (7),

$$J_{\boldsymbol{\mu}}(\mathbf{x}) = \lim_{K\to\infty} \frac{1}{K}\mathbb{E}\left\{ \sum_{k=0}^{\widetilde{K}-1} g\left(\mathbf{x}_k, \boldsymbol{\mu}(\mathbf{x}_k)\right) \,\middle|\, \mathbf{x}_0 = \mathbf{x} \right\} + \lim_{K\to\infty} \frac{1}{K}\mathbb{E}\left\{ \sum_{k=\widetilde{K}}^{K-1} g\left(\mathbf{x}_k, \boldsymbol{\mu}(\mathbf{x}_k)\right) \,\middle|\, \mathbf{x}_{\widetilde{K}} = \mathbf{y} \right\} \tag{11}$$

$$= 0 + \lim_{K\to\infty} \frac{1}{K}\mathbb{E}\left\{ \sum_{k=\widetilde{K}}^{K-1} g\left(\mathbf{x}_k, \boldsymbol{\mu}(\mathbf{x}_k)\right) \,\middle|\, \mathbf{x}_{\widetilde{K}} = \mathbf{y} \right\} \tag{12}$$

$$= J_{\boldsymbol{\mu}}(\mathbf{y}) \tag{13}$$

Since the expected TSTT is independent of the initial state for every stationary policy, the same is true for the optimal policy. Hence, $J^*(\mathbf{x}) = J^*(\mathbf{y})$, $\forall\, \mathbf{x}, \mathbf{y} \in S$. ■

Thus, Proposition 1 implies that the optimal expected TSTT is independent of the initial conditions, i.e., state of the system on day 0.

Proposition 2 (Bellman's equation)**.** *Suppose $\lambda^* = J^*(\mathbf{x})$. Then exists $h^*(\mathbf{x})\, \forall\, \mathbf{x} \in S$ (not necessarily unique) that satisfies the following Bellman's equation*

$$\lambda^* + h^*(\mathbf{x}) = \min_{\mathbf{u}\in U(\mathbf{x})} \left\{ g(\mathbf{x}, \mathbf{u}) + \sum_{\mathbf{y}\in S} p_{\mathbf{xy}}(\mathbf{u}) h^*(\mathbf{y}) \right\} \forall\, \mathbf{x} \in S \tag{14}$$

Also, if some λ and a vector of h's satisfy (14)*, then λ is the optimal average cost per stage. Further, a policy $\boldsymbol{\mu}^*(\mathbf{x})$ defined as follows is optimal*

$$\boldsymbol{\mu}^*(\mathbf{x}) \in \operatorname*{argmin}_{\mathbf{u}\in U(\mathbf{x})} \left\{ g(\mathbf{x}, \mathbf{u}) + \sum_{\mathbf{y}\in S} p_{\mathbf{xy}}(\mathbf{u}) h^*(\mathbf{y}) \right\} \forall\, \mathbf{x} \in S \tag{15}$$

Proof. See Bertsekas (2007). ■

Since the problem has a finite state space and a finite action space, the optimal J values and policies can be computed using value iteration, policy iteration or linear programming (LP). The value iteration method updates J's in the following manner and $\lambda^* = J^*(\mathbf{x})$ is obtained by evaluating $\lim_{k\to\infty} \frac{J_k(\mathbf{x})}{k}$ (k denotes the iteration number).

$$J_{k+1}(\mathbf{x}) = \min_{\mathbf{u}\in U(\mathbf{x})} \left\{ g(\mathbf{x}, \mathbf{u}) + \sum_{\mathbf{y}\in S} p_{\mathbf{xy}}(\mathbf{u}) J_k(\mathbf{y}) \right\} \forall\, \mathbf{x} \in S \tag{16}$$

where $J_0(\mathbf{x})$ can be initialized to any arbitrary value for all $\mathbf{x} \in S$. However, such an iterative procedure can lead to numerical instability as $J_k(\mathbf{x}) \to \infty$. This issue is typically avoided using relative value iteration in which we define a *differential cost vector* h_k as $h_k(\mathbf{x}) = J_k(\mathbf{x}) - J_k(\mathbf{s})\, \forall\, \mathbf{x} \in S$, where $\mathbf{s}$ is an arbitrary state in S. Hence for all $\mathbf{x} \in S$,

$$h_{k+1}(\mathbf{x}) = J_{k+1}(\mathbf{x}) - J_{k+1}(\mathbf{s}) \tag{17}$$

$$= \min_{\mathbf{u}\in U(\mathbf{x})} \left\{ g(\mathbf{x}, \mathbf{u}) + \sum_{\mathbf{y}\in S} p_{\mathbf{xy}}(\mathbf{u}) J_k(\mathbf{y}) \right\} - \min_{\mathbf{u}\in U(\mathbf{s})} \left\{ g(\mathbf{s}, \mathbf{u}) + \sum_{\mathbf{y}\in S} p_{\mathbf{sy}}(\mathbf{u}) J_k(\mathbf{y}) \right\} \tag{18}$$

$$= \min_{\mathbf{u}\in U(\mathbf{x})} \left\{ g(\mathbf{x},\mathbf{u}) + \sum_{\mathbf{y}\in S} p_{\mathbf{xy}}(\mathbf{u}) h_k(\mathbf{y}) \right\} - \min_{\mathbf{u}\in U(\mathbf{s})} \left\{ g(\mathbf{s},\mathbf{u}) + \sum_{\mathbf{y}\in S} p_{\mathbf{sy}}(\mathbf{u}) h_k(\mathbf{y}) \right\} \tag{19}$$

The iterates generated by (19) and λ_{k+1} defined according to (20) converge and satisfy the Bellman's equation as defined in Proposition 2 (Puterman, 2005; Bertsekas, 2007).

$$\lambda_{k+1} = \min_{\mathbf{u}\in U(\mathbf{s})} \left\{ g(\mathbf{s},\mathbf{u}) + \sum_{\mathbf{y}\in S} p_{\mathbf{sy}}(\mathbf{u}) h_{k+1}(\mathbf{y}) \right\} \tag{20}$$

The pseudocode for relative value iteration is summarized in Algorithm 1. Let $\epsilon > 0$ denote the required level of convergence. Suppose that $M > \epsilon$ and $\mathrm{sp}(\cdot)$ represents the span semi-norm which is defined as $\mathrm{sp}(h) = \max_{\mathbf{x}\in S} h(\mathbf{x}) - \min_{\mathbf{x}\in S} h(\mathbf{x})$. The span semi-norm is used to compute the difference between the upper and lower bounds of the optimal expected TSTT λ^*.

Algorithm 1 *Pseudocode for relative value iteration*

Step 1:
Initialize $h_0(\mathbf{x}) \,\forall\, \mathbf{x} \in S$ to any arbitrary values
$error \leftarrow M$
$k \leftarrow 0$

Step 2:
while $error > \epsilon$ **do**
  for each $\mathbf{x} \in S$ **do**
    $(Th_k)(\mathbf{x}) \leftarrow \min_{\mathbf{u}\in U(\mathbf{x})} \left\{ g(\mathbf{x},\mathbf{u}) + \sum_{\mathbf{y}\in S} p_{\mathbf{xy}}(\mathbf{u}) h_k(\mathbf{y}) \right\}$
    $h_{k+1}(\mathbf{x}) \leftarrow (Th_k)(\mathbf{x}) - (Th_k)(\mathbf{s})$
  end for
  if $k \geq 1$ **then** $error \leftarrow \mathrm{sp}(Th_k - Th_{k-1})$
  $k \leftarrow k + 1$
end while

Step 3: Choose $\boldsymbol{\mu}^*(\mathbf{x}) \in \arg\min_{\mathbf{u}\in U(\mathbf{x})} \left\{ g(\mathbf{x},\mathbf{u}) + \sum_{\mathbf{y}\in S} p_{\mathbf{xy}}(\mathbf{u})(Th_{k-1})(\mathbf{y}) \right\} \forall\, \mathbf{x} \in S$

4 Approximate dynamic programming – State space aggregation

The model formulated in the previous section, while being theoretically appealing, may not be suited for practical implementation especially when there are a large number of travelers or if there are many routes to choose from. For instance if 1000 travelers make route choices each day in a network with 10 routes, the size of the state space is equal to $\binom{1000+10-1}{1000} \approx 10^{21}$. The problem further gets compounded when we extend the model to multiple OD pairs. In this section, we address this issue by developing approximation methods that involve state space aggregation[1].

[1]State space aggregation was preferred over other approximate dynamic programming methods such as rollout algorithms (Bertsekas et al., 1997) and approximate linear programming (de Farias and Van Roy, 2006) because it avoids the enumeration of states and the computation of multinomially distributed transition probabilities.

When dealing with networks with a large number of travelers, several states may not be significantly different from each other. For instance, if there are a 1000 travelers in a network with two parallel links, the states (1000,0) and (999,1) are likely to be indistinguishable both in terms of the associated travel times and the optimal policies. This motivates us to develop approximate dynamic programming methods by aggregating states in order to reduce the computational times. Thus, we need to address the following questions: (1) how should states be aggregated? and (2) what are the transition probabilities between states in the aggregated system?

A simple attempt to aggregate states may be made using intervals of some chosen width. For instance, in the network with two parallel links, we may group states for which the flow on one of the links (say the top link) is between 0 and 10, 11 and 20 and so on. For any such aggregation/partition of the set S, transition probabilities between aggregated states may be computed by adding the transition probabilities between every pair of original states within two aggregated states. Although we save on the time required to compute the optimal policy, we would still have to enumerate all states and calculate the transition probabilities associated with states in the original state space as given by the expressions derived in (4).

Alternately, we can exploit the fact that for large n, a multinomial distributed random variable may be approximated to have a multivariate normal distribution (Sheffi, 1985). In order to do so, we first assume that the state space is continuous by supposing that travelers are infinitesimally divisible (non-atomic). Let $\mathbf{y}$ represent a vector of random variables (path flows) when action $\mathbf{u}$ is taken in state $\mathbf{x}$. In (1) we saw that $\mathbf{y}|\mathbf{x},\mathbf{u}$ is multinomially distributed. When n is large, we can approximate it with the multivariate normal $\mathbf{y}|\mathbf{x},\mathbf{u} \sim \mathcal{N}\left(\boldsymbol{\alpha}(\mathbf{x},\mathbf{u}), \boldsymbol{\Sigma}(\mathbf{x},\mathbf{u})\right)$, where $\boldsymbol{\alpha}(\mathbf{x},\mathbf{u}) = \Big(nq_1(\mathbf{x},\mathbf{u}), nq_2(\mathbf{x},\mathbf{u}), \ldots, nq_r(\mathbf{x},\mathbf{u})\Big)$ and $\boldsymbol{\Sigma}(\mathbf{x},\mathbf{u}) = [\Sigma_{ij}(\mathbf{x},\mathbf{u})]$ is the covariance matrix with elements given by

$$\Sigma_{ij}(\mathbf{x},\mathbf{u}) = \begin{cases} -nq_i(\mathbf{x},\mathbf{u})q_j(\mathbf{x},\mathbf{u}) & \text{if } i \neq j \\ nq_i(\mathbf{x},\mathbf{u})(1-q_i(\mathbf{x},\mathbf{u})) & \text{otherwise} \end{cases} \tag{21}$$

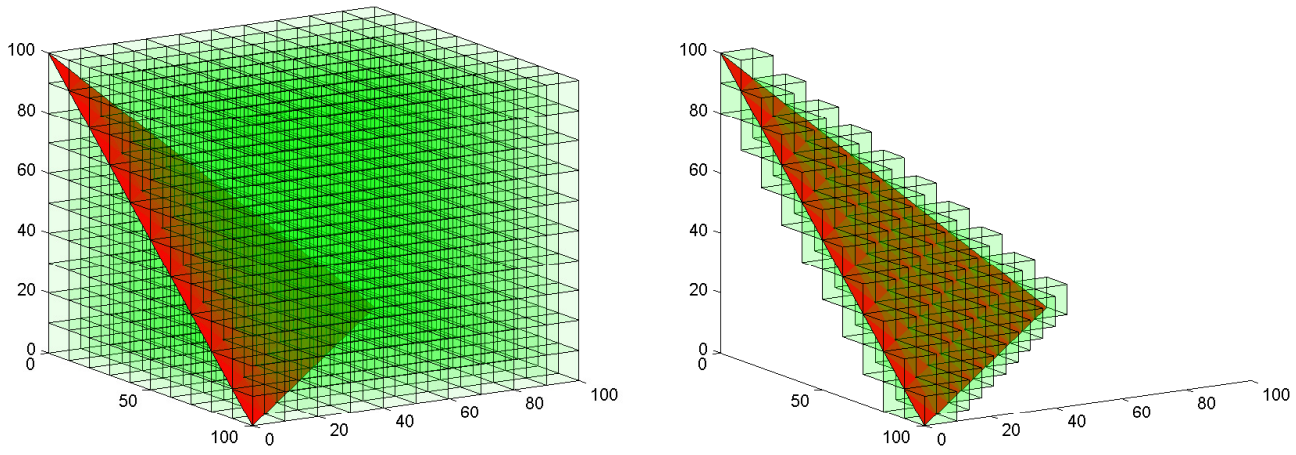
The density function of $\mathbf{y}$ is given by

$$f(\mathbf{y}|\mathbf{x},\mathbf{u}) = \frac{1}{\sqrt{(2\pi)^r \det \boldsymbol{\Sigma}(\mathbf{x},\mathbf{u})}} \exp\left(-\frac{1}{2}\left(\mathbf{y}-\boldsymbol{\alpha}(\mathbf{x},\mathbf{u})\right)^T \boldsymbol{\Sigma}(\mathbf{x},\mathbf{u})^{-1}\left(\mathbf{y}-\boldsymbol{\alpha}(\mathbf{x},\mathbf{u})\right)\right) \tag{22}$$

State space

The theory of infinite horizon MDPs is well established for problems with finite state and action spaces. In order to take advantage of existing methods to solve them, we construct a finite number of states from a continuous state space by generalizing the idea of aggregating states using intervals. Let us first define the set $\mathcal{I} = \left\{[0, \frac{n}{\delta}], [\frac{n}{\delta}, \frac{2n}{\delta}], \ldots, [\frac{(\delta-1)n}{\delta}, n]\right\}$. Notice that $\mathcal{I}^r$ is the set of all hypercubes formed by dividing the flow on each route into δ intervals. We then consider the space $\mathcal{S} = \{\mathcal{X} \in \mathcal{I}^r : |\mathcal{X} \cap \{\mathbf{x} \in [0,n]^r : \sum_{i=1}^r x_i = n\}| > 1\}$, where $|\cdot|$ represents the cardinality of a set. Figure 3 helps visualize this construct. Suppose there are 100 travelers and three routes, and the flows on each route are represented on the three axes. Assume that we divide the each axis into 10 intervals. This divides the space $[0,100]^3$ into 1000 hypercubes as shown in Figure 3a. We then pick only those hypercubes which intersect the set of feasible flows (i.e, the simplex

$x_1 + x_2 + x_3 = 100$) at more than one point, which gives the 100 hypercubes in Figure 3b. We exclude the hypercubes that intersect the simplex at exactly one point as we can always find another hypercube belonging to $\mathcal{S}$ that contains the point.



(a) Set of all hypercubes $\mathcal{I}^r$ *(b) Hypercubes that intersect the simplex*

Figure 3: State space for the approximate methods

Let the state space for the approximate method be $\mathcal{S}$. For any state $\mathcal{X} \in \mathcal{S}$, let $\mathcal{X}_c \in \mathbb{R}^r$ be the center of the hypercube $\mathcal{X}$. We evaluate the properties of the state $\mathcal{X}$ such as the TSTT at this point. Notice that the point $\mathcal{X}_c$ may or may not satisfy the flow conservation constraint depending on the choice of r and δ. However, when we consider a sufficiently large number of intervals (δ), $\mathcal{X}_c$ may be assumed to be close enough to the simplex so that the errors in approximating the TSTT are small. We now define the remaining components of the MDP.

Action space

The action space for the approximate MDP at each state is same as before, i.e., $U(\mathcal{X}) = \{\tau_1, \tau_2, \ldots, \tau_l\}^r$.

Transition probabilities

Let the transition probabilities of moving from state $\mathcal{X}$ to $\mathcal{Y}$ under action $\mathbf{u}$ be denoted as $p_{\mathcal{X}\mathcal{Y}}(\mathbf{u})$. The transition probabilities may be approximated using the cumulative densities of the multivariate normal but this may lead to values that do not add up to 1. Hence, we first define $p'_{\mathcal{X}\mathcal{Y}}(\mathbf{u})$ as

$$p'_{\mathcal{X}\mathcal{Y}}(\mathbf{u}) = \int_{\mathbf{y}\in\mathcal{Y}} \frac{1}{\sqrt{(2\pi)^r \det \boldsymbol{\Sigma}(\mathcal{X}_c, \mathbf{u})}} \exp\left(-\frac{1}{2}(\mathbf{y} - \boldsymbol{\alpha}(\mathcal{X}_c, \mathbf{u}))^T \boldsymbol{\Sigma}(\mathcal{X}_c, \mathbf{u})^{-1} (\mathbf{y} - \boldsymbol{\alpha}(\mathcal{X}_c, \mathbf{u}))\right) d\mathbf{y} \tag{23}$$

where $\alpha(\mathcal{X}_c, \mathbf{u})$ and $\Sigma(\mathcal{X}_c, \mathbf{u})$ are defined as mentioned earlier. Next, we normalize these values by setting $p_{\mathcal{X}\mathcal{Y}}(\mathbf{u}) = p'_{\mathcal{X}\mathcal{Y}}(\mathbf{u}) / \sum_{\mathcal{Y}'\in\mathcal{S}} p'_{\mathcal{X}\mathcal{Y}'}(\mathbf{u})$.

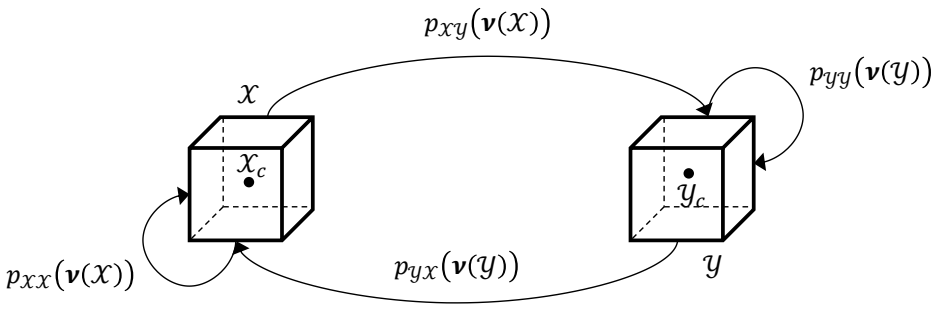


Figure 4: Transitions between aggregated states

Costs

The cost incurred in choosing $\mathbf{u}$ in state $\mathcal{X}$, $g(\mathcal{X}, \mathbf{u})$, is defined as $g(\mathcal{X}, \mathbf{u}) = \sum_{\mathcal{Y} \in \mathcal{S}} p_{\mathcal{X}\mathcal{Y}}(\mathbf{u}) \text{TSTT}(\mathcal{Y}_c)$.

The objective for the approximate MDP is defined similarly as in Section 3.2. Let $J_{\boldsymbol{\nu}}(\mathcal{X})$ be the average cost per stage for policy $\boldsymbol{\nu}$ when the system starts at state $\mathcal{X}$. Assuming that $\mathcal{X}_k$ represents a state on the k^{th} day in the aggregated system,

$$J_{\boldsymbol{\nu}}(\mathcal{X}) = \lim_{K \to \infty} \frac{1}{K} \mathbb{E} \left\{ \sum_{k=0}^{K-1} g\left(\mathcal{X}_k, \boldsymbol{\nu}(\mathcal{X}_k)\right) \,\middle|\, \mathcal{X}_0 = \mathcal{X} \right\} \tag{24}$$

where $\big(\boldsymbol{\nu}(\mathcal{X})\big)_{\mathcal{X} \in \mathcal{S}}$ specifies the action $\boldsymbol{\nu}(\mathcal{X}) \in U(\mathcal{X})$ to be taken when the system is in state $\mathcal{X}$. Let the optimal policy be denoted by $\boldsymbol{\nu}^*$, i.e., $J^*(\mathcal{X}) \equiv J_{\boldsymbol{\nu}^*}(\mathcal{X}) = \min_{\boldsymbol{\nu} \in \Phi} J_{\boldsymbol{\nu}}(\mathcal{X})$, where Φ is the set of all admissible policies. Since the state and action spaces are finite, Algorithm 1 can be applied to solve the approximate MDP.

Let $\Gamma : S \to \mathcal{S}$ be a mapping that gives the aggregated state to which a state in the original state space belongs (ties are broken arbitrarily). Then an approximate optimal policy for the original MDP can be defined as $\big(\boldsymbol{\mu}(\mathbf{x})\big)_{\mathbf{x} \in S}$, where $\boldsymbol{\mu}(\mathbf{x}) = \boldsymbol{\nu}^*\big(\Gamma(\mathbf{x})\big)$.

5 Demonstration

The method described in the previous section provides a policy that is optimal for the approximate MDP and an immediate question of interest is if it is close to the optimal policy for the original MDP. One possible way to answer this question is by tracking the errors involved. However, this is extremely difficult as we are making several approximations to the transition probabilities and in aggregating states. Instead, in this paper, we resort to numerical experimentation to make claims about the approximate policy. While computing the optimal expected TSTT for large n is difficult, for small values of n, we can use relative value iteration or other methods to exactly compute the optimal expected TSTT which can thus be used to ascertain how far we are from the optimal.

For any n, clearly the expected TSTT of the no-tolls case (or the do-nothing option) gives an upper bound to the optimal TSTT. Calculating the expected TSTT of a given policy is relatively easy and can be done

by estimating the steady state distribution of the Markov chain under that policy or by simulation. Thus, we can estimate the expected TSTT under the approximate policy $\boldsymbol{\nu}^*(\Gamma(\mathbf{x}))$ and check if is an improvement over the no-tolls option, i.e., if it provides a better upper bound.

As n increases, the quality of approximations made to the state space and transition probabilities improves and depending on the available computing power, one can pick larger δ to develop finer partition schemes. Hence, we claim that this empirical line of analysis is proof enough that this method may be applied to problems with large state spaces.

For the numerical results presented in this paper, we consider the network in Figure 5. Each traveler has three routes 1-2-4, 1-2-3-4, and 1-3-4. The link level travel times are assumed to be a function of the link flows and are shown in the figure. We assume that the set of possible tolls on each route can be enumerated as $\{0, 2, 4, \ldots, 8\}$, i.e., the action space for each state is $\{0, 2, 4, \ldots, 8\}^3$. The methods described were implemented in C++ (using the g++ compiler with -O2 optimization flags) on a Linux machine with 24 core Intel Xeon processors (3.33 GHz) and 12 MB cache. The cumulative densities for the multivariate normal distribution were obtained using a Fortran function available at `http://www.math.wsu.edu/faculty/genz/software/software.html`. The termination criterion for relative value iteration was set to 1E-07. The value of the dispersion parameter θ was fixed at 0.1. Most of the code was implemented in a parallel environment using OpenMP except for function calls to the Fortran code.

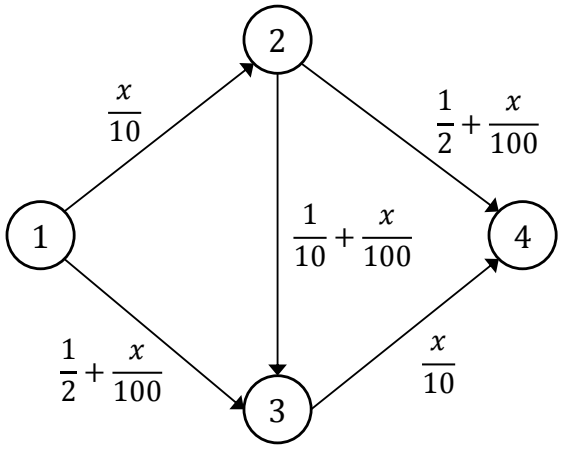


Figure 5: Network used to test the approximations

Solution quality

Table 1 compares the expected TSTT of the optimal and approximate policies for different number of travelers. It can be observed that the approximate policies perform better than the no-tolls option and the quality of approximation gets better with increase in δ. In all four cases, the approximate policies were found to be optimal or nearly optimal for large δ.

Table 1: Comparison on expected TSTT of policies

n	*Optimal expected TSTT*	*Expected TSTT of approximate policy* $\delta = 5$	$\delta = 10$	$\delta = 20$	*Expected TSTT for no-tolls*
50	200.012	200.012	200.012	200.012	233.966
100	720.506	746.378	720.520	720.510	830.267
150	1532.220	1658.040	1532.930	1532.530	1730.410
200	2618.180	2881.280	2619.450	2618.750	2932.050

Computational performance

Table 2 indicates the wall-clock time in seconds for the steps involved in solving the exact and approximate MDPs. As mentioned earlier, the algorithms were implemented in a parallel environment with 24 cores except for the computation of transition probabilities of the approximate MDPs. Since these probabilities can be computed independently of each other, one can expect near linear speedup if implemented in a parallel manner. Notice that for $n = 100$, the value iteration for the exact method takes nearly 50 minutes and on the other hand the approximate methods provide near optimal solutions within a few seconds. For the exact MDP, when $n = 150$ and $n = 200$, the memory requirements for storing the transition probabilities exceeded available computing resources and hence they were not stored but were recomputed within each value iteration step. The run times for these instances were around 3 to 5 hours and have been left out of Table 2 in order to provide a fair comparison.

The results appear promising and for problems with much larger n, we may choose a value of δ according to the available computational resources. As the quality of the approximations get better with increase in n, the approximate policies can be expected to perform better than the no-toll policy.

Table 2: Wall-clock times (in seconds) for exact and approximate methods

Number of Travelers $\rightarrow$		50	100	150	200
Exact MDP	*No. of states*	1326	5151	11476	20301
	State space	0.023	0.193	0.561	1.316
	Trans prob	5.370	229.614	-	-
	Value itn	2.597	3056.490		
Approx MDP ($\delta = 5$)	*No. of states*	25	25	25	25
	State space	5.80E-05	7.40E-05	5.70E-05	6.00E-05
	Trans prob	293.713	198.249	145.295	103.442
	Value itn	0.210	1.558	0.013	0.040
Approx MDP ($\delta = 10$)	*No. of states*	100	100	100	100
	State space	5.13E-04	3.69E-04	3.36E-04	3.04E-04
	Trans prob	1351.90	1785.57	1461.96	1163.03
	Value itn	0.189	0.346	0.388	0.016
Approx MDP ($\delta = 20$)	*No. of states*	400	400	400	400
	State space	3.79E-03	3.77E-03	2.60E-03	2.53E-03
	Trans prob	9538.21	9505.30	9606.38	9733.46
	Value itn	0.377	0.451	0.243	0.425

Mixing times

The Markov chain associated with any given policy is irreducible and aperiodic and hence has a steady

state distribution. However, it takes a certain amount of time for the Markov chain to get "close" to its the stationary distribution. While, this is not a major concern for the average cost MDP, from a theoretical perspective, since we let $k \to \infty$, it would be useful to know how long it takes the Markov chain to reach its stationary distribution from a practical standpoint.

This question can be answered by analyzing the mixing time of the Markov chain associated with the optimal policy. In order to do so, a few definitions are in order. Let $\|\cdot\|_{TV}$ represent the total variation distance, which is a measure of the distance between two probability distributions. For any two probability density functions π and π', the total variation distance is defined as $\|\pi - \pi'\|_{TV} = \max_{A \subset S} |\pi(A) - \pi'(A)|$. Further, if S is discrete, it can be shown that $\|\pi - \pi'\|_{TV} = \frac{1}{2} \max_{\mathbf{x} \in S} |\pi(\mathbf{x}) - \pi'(\mathbf{x})|$ (Levin et al., 2009).

Now let P represent the transition probability matrix associated with the optimal policy $\boldsymbol{\mu}^*$, i.e., $P(\mathbf{x}, \mathbf{y}) = p_{\mathbf{xy}}(\boldsymbol{\mu}^*(\mathbf{x}))$ and let $\pi(\mathbf{x})$ denote the steady state probability of observing state $\mathbf{x}$. We define the maximum variation distance $d(k)$ between the rows of P and π after k steps as follows:

$$d(k) = \max_{\mathbf{x} \in S} \|P^k(\mathbf{x}, \cdot) - \pi(\cdot)\|_{TV} \tag{25}$$

Since a steady state distribution exists, $d(k) \to 0$ as $k \to \infty$. Further, for irreducible and aperiodic Markov chains, the rate at which $d(k)$ shrinks to zero is exponential and is bounded below by $\frac{1}{2}(1-\gamma)^k$, where γ is the absolute spectral gap which equals one minus the second largest magnitude eigenvalue (Montenegro and Tetali, 2006). Thus, the higher the value of gamma, the faster the rate at which the Markov chain converges to its steady state distribution. Another way to analyze the mixing times is by observing the least number of time steps before $d(k)$ falls below an arbitrary threshold ϵ.

$$t_{mix}(\epsilon) = \min\{k : d(k) \leq \epsilon\} \tag{26}$$

Table 3 shows the spectral gap and mixing times for the Markov chains associated with the optimal policy for the four cases tested earlier. The spectral gap is not close to 0 and hence the Markov chains mix fairly quickly.

Table 3: Spectral gap and mixing times for different problem instances

Number of travelers (n)	*Spectral gap* (γ)	*Mixing time* ($t_{mix}(0.01)$)
50	0.767	3
100	0.530	6
150	0.491	7
200	0.560	6

Finite Markov chains are often known to abruptly convergence to their stationary distributions. This feature, also known as the cutoff phenomena (Diaconis, 1996; Chen, 2006), results in a sudden drop in the $d(k)$ values. Figure 6 shows the maximum variation distance across consecutive days for different problem instances. While no evidence of the cutoff phenomenon was found, it may be interesting to see if such phenomena occur in problems with larger state spaces. We, however, observed that the mixing times of

Markov chains increase with increase in the size of the state space.

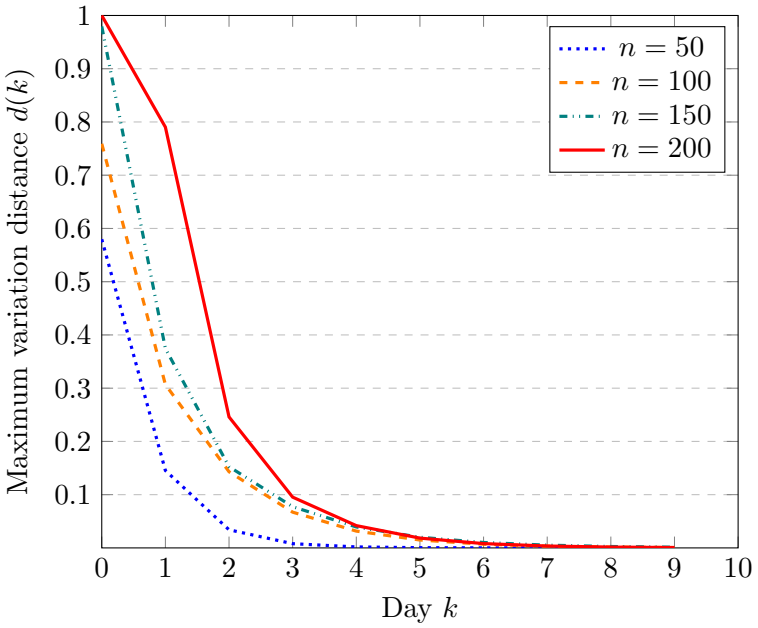


Figure 6: Variation distance of Markov chains associated with the optimal policy

6 Discussion

In this section, we first present some other objectives that may be of interest to a system manager or a private tolling agency. We then conclude by summarizing the methods proposed in this paper and discuss potential future research ideas.

6.1 Variants

6.1.1 Other network-wide objectives

By defining the average costs/rewards for a state-action pair differently we can find dynamic pricing policies that optimize other objectives.

Maximizing the probability of convergence to a target state: Suppose we wish to increase the probability of finding the system in a particular state in the long run. We will henceforth refer to this state as the *target state*. While a SO flow solution is an obvious choice for the target state, one could think of other target states based on other network wide objectives such as emissions. Also, in the presence of multiple NE solutions, one equilibrium may be favored over another and be chosen as the target state. In order to achieve this objective, we define the *rewards* (instead of costs) as follows:

$$g(\mathbf{x}, \mathbf{u}) = \begin{cases} 1 \text{ if } \mathbf{x} \text{ is the target state} \\ 0 \text{ otherwise} \end{cases} \tag{27}$$

Thus, every time the system leaves the target state we receive a reward of 1 unit and therefore the long run probability of being in the target state is the average reward per stage. Instead, if we want to increase

the probability of finding the system in set of states (i.e., we have a set of target states), we could just set $g(\mathbf{x}, \mathbf{u}) = 1$ for all $\mathbf{x}$ that belong to such a set and 0 otherwise.

One can think of extensions to these problems in which tolls are disallowed at the target state, i.e., if the network is at the target state on a particular day, then the system manager may choose to not collect any tolls for the next day. This mechanism has the same flavor as that of punishment strategies in repeated games that are used to force players to cooperate. This feature can easily be modeled by setting the action space at the target state to the empty set. The objective is likely to be lower when tolls are disallowed. Yet, by pricing all states except the target state, this formulation can lead to an increase in the probability of reaching the target state.

Minimizing the expected deviation from TSTT($\mathbf{x}_{SO}$): Suppose we want to minimize the deviation from the TSTT of the SO state. This objective could be useful in the context of improving travel time reliability as it can help reduce the variance in travel times and may be achieved by defining the stage costs as follows:

$$g(\mathbf{x}, \mathbf{u}) = (\text{TSTT}(\mathbf{x}) - \text{TSTT}(\mathbf{x}_{SO}))^2 \, \forall \, \mathbf{u} \in U(\mathbf{x}) \tag{28}$$

6.1.2 Incentives and revenue maximization

Assume that the system manager can incentivize travelers in addition to collecting tolls. Suppose that we model incentives as negative tolls. The optimal policy in such cases may require the system manager to pay something to travelers on an average. We can avoid this by adding side constraints (32) to the LP model (see Bertsekas (2007) for details) for the average cost MDP as shown below. Let $b(\mathbf{x}, \mathbf{u})$ be the expected revenue/cost for the system manager when $\mathbf{u}$ is chosen at state $\mathbf{x}$. The idea behind adding the side constraints is similar to budget balance mechanisms that are studied in mechanism design in which payment schemes that ensure zero net payments to all players are sought.

$$\lambda^* = \min \sum_{\mathbf{x} \in S} \sum_{\mathbf{u} \in U(\mathbf{x})} d(\mathbf{x}, \mathbf{u}) g(\mathbf{x}, \mathbf{u}) \tag{29}$$

$$\text{s.t.} \sum_{\mathbf{u} \in U(\mathbf{y})} d(\mathbf{y}, \mathbf{u}) = \sum_{\mathbf{x} \in S} \sum_{\mathbf{u} \in U(\mathbf{x})} d(\mathbf{x}, \mathbf{u}) p_{\mathbf{xy}}(\mathbf{u}) \;\; \forall \, \mathbf{y} \in S \tag{30}$$

$$\sum_{\mathbf{x} \in S} \sum_{\mathbf{u} \in U(\mathbf{x})} d(\mathbf{x}, \mathbf{u}) = 1 \tag{31}$$

$$\sum_{\mathbf{x} \in S} \sum_{\mathbf{u} \in U(\mathbf{x})} d(\mathbf{x}, \mathbf{u}) b(\mathbf{x}, \mathbf{u}) \geq 0 \tag{32}$$

$$d(\mathbf{x}, \mathbf{u}) \geq 0 \;\; \forall \, \mathbf{x} \in S, \mathbf{u} \in U(\mathbf{x}) \tag{33}$$

In the above LP model, (30) represents the balance equations and (31) is the normalization constraint. The optimal values of $d^*(\mathbf{x}, \mathbf{u})$ can be used to construct the steady state distribution and the optimal policy. More precisely, for ergodic MDPs, for every state $\mathbf{x} \in S$, there exists exactly one $\mathbf{u} \in U(\mathbf{x})$ for which $d^*(\mathbf{x}, \mathbf{u}) > 0$. Setting $\boldsymbol{\mu}^*(x)$ to $\mathbf{u} \in U(\mathbf{x})$ for which $d^*(\mathbf{x}, \mathbf{u}) > 0$ gives the optimal policy. Further,

$d^*(\mathbf{x}, \boldsymbol{\mu}^*(x))$ denotes the steady state probability of finding the system in state $\mathbf{x}$ under the optimal policy. Hence, the objective represents the expected TSTT and the left hand side of constraint (32) computes the expected revenue/cost.

However LP models are well suited for problems with small state and action spaces. We may therefore use the approximation methods developed in Section 4 and formulate the LP using the aggregated state space. Also, since travelers do not perceive incentives and tolls the same way, one can use prospect theory (Kahneman and Tversky, 1979) to distinguish between these. In addition, if the system manager wishes to achieve a certain target revenue (which could in turn be used for maintaining the tolling infrastructure), we can set the right hand side of constraint (32) to the expected profit or target value.

6.2 Conclusions

In this paper, we developed a dynamic day-to-day pricing model that can help a system manager minimize the expected TSTT. Specifically, we formulated the problem as an infinite horizon average cost MDP which provides stationary policies that are a function of the state of the system. Since practical problems involve a large number of travelers and exponential state spaces, we proposed approximate solution methods and performed a few numerical experiments to test their quality.

The results indicate that (1) for large number of travelers, the approximate policies obtained by state space aggregation methods result in a significant reduction of expected TSTT compared to the no-toll case and (2) the computation times for obtaining an approximate optimal policy are reduced to a tractable degree after aggregating states.

The findings in this paper call for exploring several scenarios which relax the assumptions made. For instance, in practice, users may not in respond to tolls the same way because they may have different monetary values for time. Further, the route choice mechanism for all travelers may not be well captured using a single logit or probit choice model. In such cases, we could use reinforcement learning approaches to learn how travelers respond to congestion and pricing, and compute policies in an online manner. Also, it would be interesting to extend the current models to problems with elastic demand.

Acknowledgments

This research was supported by the Data-Supported Transportation Operations and Planning (D-STOP), University Transportation Center. The authors are solely responsible for any errors or omissions in this research paper. The authors would like to thank Kai Yin and John Hasenbein for useful discussions on this topic and two anonymous reviewers for providing several insightful comments.

References

Hillel Bar-Gera. Origin-based algorithm for the traffic assignment problem. *Transportation Science*, 36(4): 398–417, 2002.

Dimitri P. Bertsekas. *Dynamic Programming and Optimal Control*, volume 2. Athena Scientific, 4th edition edition, 2007.

Dimitri P. Bertsekas, John N. Tsitsiklis, and Cynara Wu. Rollout algorithms for combinatorial optimization. *Journal of Heuristics*, 3(3):245–262, 1997.

Lawrence Blume. Population games. Game Theory and Information 9607001, EconWPA, July 1996.

George W. Brown. Iterative solution of games by fictitious play. In T.C. Koopmans, editor, *Activity Analysis of Production and Allocation*, pages 374–376. Wiley, New York, 1951.

George W. Brown and John Von Neumann. Solutions of games by differential equations. Technical report, DTIC Document, 1950.

Giulio E. Cantarella and Ennio Cascetta. Dynamic processes and equilibrium in transportation networks: towards a unifying theory. *Transportation Science*, 29(4):305–329, 1995.

Ennio Cascetta. A stochastic process approach to the analysis of temporal dynamics in transportation networks. *Transportation Research Part B: Methodological*, 23(1):1 – 17, 1989.

Ennio Cascetta and Giulio E. Cantarella. A day-to-day and within-day dynamic stochastic assignment model. *Transportation Research Part A: General*, 25(5):277 – 291, 1991.

Guan-Yu Chen. *The cutoff phenomenon for finite Markov chains*. PhD thesis, Cornell University, 2006.

Carlos F. Daganzo and Yosef Sheffi. On stochastic models of traffic assignment. *Transportation Science*, 11(3):253–274, 1977.

Gary A. Davis and Nancy L. Nihan. Large population approximations of a general stochastic traffic assignment model. *Operations Research*, 41(1):169–178, 1993.

Daniela Pucci de Farias and Benjamin Van Roy. A cost-shaping linear program for average-cost approximate dynamic programming with performance guarantees. *Mathematics of Operations Research*, 31(3):597–620, 2006.

Persi Diaconis. The cutoff phenomenon in finite markov chains. *Proceedings of the National Academy of Sciences*, 93(4):1659–1664, 1996.

Robert B. Dial. A probabilistic multipath traffic assignment model which obviates path enumeration. *Transportation Research*, 5:83–111, 1971.

Robert B. Dial. A path-based user-equilibrium traffic assignment algorithm that obviates path storage and enumeration. *Transportation Research Part B*, 40(10):917–936, 2006.

Farhad Farokhi and Karl H. Johansson. A piecewise-constant congestion taxing policy for repeated routing games. *Transportation Research Part B: Methodological*, 78(0):123 – 143, 2015.

Terry L. Friesz, David Bernstein, Nihal J. Mehta, Roger L. Tobin, and Saiid Ganjalizadeh. Day-to-day dynamic network disequilibria and idealized traveler information systems. *Operations Research*, 42(6): 1120–1136, 1994.

Terry L. Friesz, David Bernstein, and Niko Kydes. Dynamic congestion pricing in disequilibrium. *Networks and Spatial Economics*, 4(2):181–202, 2004. ISSN 1566-113X.

Ren-Yong Guo, Hai Yang, and Hai-Jun Huang. A discrete rational adjustment process of link flows in traffic networks. *Transportation Research Part C: Emerging Technologies*, 34(0):121 – 137, 2013.

Ren-Yong Guo, Hai Yang, Hai-Jun Huang, and Zhijia Tan. Link-based day-to-day network traffic dynamics and equilibria. *Transportation Research Part B: Methodological*, 71(0):248 – 260, 2015. ISSN 0191-2615.

Lanshan Han and Lili Du. On a link-based day-to-day traffic assignment model. *Transportation Research Part B: Methodological*, 46(1):72 – 84, 2012. ISSN 0191-2615.

Martin L. Hazelton and David P. Watling. Computation of equilibrium distributions of markov traffic-assignment models. *Transportation Science*, 38(3):331–342, 2004.

Xiaozheng He, Xiaolei Guo, and Henry X. Liu. A link-based day-to-day traffic assignment model. *Transportation Research Part B: Methodological*, 44(4):597 – 608, 2010.

R. Jayakrishnan, Wei T. Tsai, Joseph N. Prashker, and Subodh Rajadhyaksha. A faster path-based algorithm for traffic assignment. *Transportation Research Record*, 1443:75–83, 1994.

Dusica Joksimovic, Michiel C. J. Bliemer, and Piet H. L. Bovy. Optimal toll design problem in dynamic traffic networks with joint route and departure time choice. *Transportation Research Record: Journal of the Transportation Research Board*, 1923(1):61–72, 2005.

Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):pp. 263–292, 1979.

Michihiro Kandori and Rafael Rob. Evolution of equilibria in the long run: A general theory and applications. *Journal of Economic Theory*, 65(2):383–414, 1995.

Michihiro Kandori, George J. Mailath, and Rafael Rob. Learning, mutation, and long run equilibria in games. *Econometrica*, 61(1):pp. 29–56, 1993.

Torbjörn Larsson and Michael Patriksson. Simplicial decomposition with disaggregated representation for the traffic assignment problem. *Transportation Science*, 26:4–17, 1992.

David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times*. American Mathematical Soc., 2009.

Maria Mitradjieva and Per O. Lindberg. The stiff is moving — conjugate direction Frank-Wolfe methods with application to traffic assignment. *Transportation Science*, 47(2):280–293, 2013.

Ravi R. Montenegro and Prasad Tetali. *Mathematical aspects of mixing times in Markov chains*. Now Publishers Inc, 2006.

Anna Nagurney and Ding Zhang. Projected dynamical systems in the formulation, stability analysis, and computation of fixed-demand traffic network equilibria. *Transportation Science*, 31(2):pp. 147–158, 1997.

John Nash. Non-cooperative games. *Annals of Mathematics*, 54(2):pp. 286–295, 1951.

Arthur C. Pigou. *The Economics of Welfare*. Macmillan and Co., London, 1920.

Martin L. Puterman. *Dynamic Programming and Optimal Control*. Wiley-Interscience, 1st edition edition, 2005.

Julia Robinson. An iterative method of solving a game. *Annals of Mathematics*, 54(2):pp. 296–301, 1951.

William H. Sandholm. *Population games and evolutionary dynamics*. MIT Press, 2010.

Yosef Sheffi. *Urban Transportation Networks*. Prentice-Hall, Englewood Cliffs, NJ, 1985.

M. J. Smith. The existence, uniqueness and stability of traffic equilibria. *Transportation Research Part B: Methodological*, 13(4):295–304, 1979.

Maynard J. Smith and GR Price. The logic of animal conflict. *Nature*, 246:15, 1973.

Tony E. Smith, Erik A. Eriksson, and Per O. Lindberg. Existence of optimal tolls under conditions of stochastic user-equilibria. In Brje Johansson and Lars-Gran Mattsson, editors, *Road Pricing: Theory, Empirical Assessment and Policy*, Transportation Research, Economics and Policy, pages 65–87. Springer Netherlands, 1995.

Peter D. Taylor and Leo B. Jonker. Evolutionary stable strategies and game dynamics. *Mathematical biosciences*, 40(1):145–156, 1978.

John G. Wardrop. Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers, Part II*, 1:352–362, 1952.

David Watling. Asymmetric problems and stochastic process models of traffic assignment. *Transportation Research Part B: Methodological*, 30(5):339 – 357, 1996.

David Watling and Martin L. Hazelton. The dynamics and equilibria of day-to-day assignment models. *Networks and Spatial Economics*, 3(3):349–370, 2003.

David P. Watling and Giulio E. Cantarella. Model representation and decision-making in an ever-changing world: The role of stochastic process models of transportation systems. *Networks and Spatial Economics*, pages 1–40, 2013.

Byung-Wook Wie and Roger L. Tobin. Dynamic congestion pricing models for general traffic networks. *Transportation Research Part B: Methodological*, 32(5):313 – 327, 1998.

Feng Xiao, Hongbo Ye, and Hai Yang. Optimal pricing of day-to-day flow dynamics. In *5th International Symposium on Dynamic Traffic Assignment*, 2014.

Fan Yang. Day-to-day dynamic optimal tolls with elastic demand. In *Transportation Research Board 87th Annual Meeting*, number 08-0305, 2008.

Fan Yang and Ding Zhang. Day-to-day stationary link flow pattern. *Transportation Research Part B: Methodological*, 43(1):119 – 126, 2009.

Hai Yang. System optimum, stochastic user equilibrium, and optimal link tolls. *Transportation Science*, 33(4):354–360, 1999.

Yafeng Yin and Yingyan Lou. Dynamic tolling strategies for managed lanes. *Journal of Transportation Engineering*, 135(2):45–52, 2009.

H. Peyton Young. The evolution of conventions. *Econometrica*, 61(1):pp. 57–84, 1993.

H. Peyton Young. *Strategic learning and its limits*, volume 2002. Oxford University Press, 2004.

Ding Zhang, Anna Nagurney, and Jiahao Wu. On the equivalence between stationary link flow patterns and traffic network equilibria. *Transportation Research Part B: Methodological*, 35(8):731 – 748, 2001.