TU Delft

Delft University of Technology

Efficient Simulation-Based Toll Optimization for Large-Scale Networks

Osorio, Carolina; Atasoy, Bilge

DOI
10.1287/trsc.2021.1043

Publication date
2021

Document Version
Final published version

Published in
Transportation Science

Citation (APA)
Osorio, C., & Atasoy, B. (2021). Efficient Simulation-Based Toll Optimization for Large-Scale Networks. *Transportation Science*, *55*(5), 1010-1024. https://doi.org/10.1287/trsc.2021.1043

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Green Open Access added to TU Delft Institutional Repository

'You share, we take care!' - Taverne project

https://www.openaccess.nl/en/you-share-we-take-care

Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.

This article was downloaded by: [154.59.124.113] On: 21 January 2022, At: 07:02
Publisher: Institute for Operations Research and the Management Sciences (INFORMS)
INFORMS is located in Maryland, USA



Transportation Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Efficient Simulation-Based Toll Optimization for Large-Scale Networks

Carolina Osorio, Bilge Atasoy

To cite this article:
Carolina Osorio, Bilge Atasoy (2021) Efficient Simulation-Based Toll Optimization for Large-Scale Networks. Transportation Science 55(5):1010-1024. https://doi.org/10.1287/trsc.2021.1043

Full terms and conditions of use: https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2021, INFORMS

Please scroll down for article—it is on subsequent pages

With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

http://pubsonline.informs.org/journal/trsc

TRANSPORTATION SCIENCE
Vol. 55, No. 5, September–October 2021, pp. 1010–1024
ISSN 0041-1655 (print), ISSN 1526-5447 (online)

Efficient Simulation-Based Toll Optimization for Large-Scale Networks

Carolina Osorio,[a] Bilge Atasoy[b]

[a] Department of Decision Sciences, HEC Montréal, 3000 Chemin de la Côte-Sainte-Catherine, Montréal H3T 2A7, Canada; [b] Department of Maritime and Transport Technology, Delft University of Technology, Delft 2628 CD, Netherlands

Contact: carolina.osorio@hec.ca, https://orcid.org/0000-0003-0979-6052 (CO); b.atasoy@tudelft.nl, https://orcid.org/0000-0002-1606-9841 (BA)

Received: October 10, 2017
Revised: June 1, 2018; February 5, 2020
Accepted: April 14, 2020
Published Online in Articles in Advance: July 30, 2021

https://doi.org/10.1287/trsc.2021.1043

Copyright: © 2021 INFORMS

Abstract. This paper proposes a simulation-based optimization technique for high-dimensional toll optimization problems of large-scale road networks. We formulate a novel analytical network model. The latter is embedded within a metamodel simulation-based optimization (SO) algorithm. It provides analytical and differentiable structural information of the underlying problem to the SO algorithm. Hence, the algorithm no longer treats the simulator as a black box. The analytical model is formulated as a system of nonlinear equations that can be efficiently evaluated with standard solvers. The dimension of the system of equations scales linearly with network size. It scales independently of the dimension of the route choice set and of link attributes such as link length. Hence, it is a scalable formulation suitable for the optimization of large-scale networks. For instance, the model is used in the case study of the paper for toll optimization of a Singapore network with more than 4,050 OD (origin-destination) pairs and 18,200 feasible routes. The corresponding analytical model is implemented as a system of 860 nonlinear equations. The analytical network model is validated based on one-dimensional toy network problems. It captures the main trends of the simulation-based objective function and, more importantly, accurately locates the global optimum for all experiments. The proposed SO approach is then used to optimize a set of 16 tolls for the network of expressways and major arterials of Singapore. The proposed method is compared with a general-purpose algorithm. The proposed method identifies good quality solutions at the very first iteration. The benchmark method identifies solutions with similar performance after 2 days of computation or similarly after more than 30 points have been simulated. The case study indicates that the analytical structural information provided to the algorithm by the analytical network model enables it to (i) identify good quality solutions fast and (ii) become robust to both the quality of the initial points and to the stochasticity of the simulator. The final solutions identified by the proposed algorithm outperform those of the benchmark method by an average of 18%.

Funding: This work was supported by Division of Civil, Mechanical, and Manufacturing Innovation and the work of C. Osorio is partially supported by the U.S. National Science Foundation [Grant 1351512].

Supplemental Material: The online appendix is available at https://doi.org/10.1287/trsc.2021.1043.

Keywords: simulation-based optimization • toll optimization • congestion pricing

1. Introduction

Transportation demand management (TDM), also referred to as travel demand management or traffic demand management, consists of strategies to reduce or redistribute travel demand in time or space, such as to improve, for instance, the efficiency or the sustainability of the transportation network. The temporal and spatial dimensions of travel (e.g., departure time, mode, route) and even the decision of whether to travel can be shaped through TDM (Saleh and Sammer 2009). TDM strategies include pricing and incentives (negative pricing). Pricing strategies are deployed more widely than incentives. Congestion pricing has been studied extensively. A recent review of congestion pricing methods is given in de Palma and Lindsey (2011). Toll optimization is one of the most studied types of congestion pricing. Both offline and online toll optimization strategies have been proposed. Offline strategies can be static (i.e., they yield a single toll for the entire time horizon of interest at a given toll location) or dynamic (i.e., they yield a time-dependent toll, such as in time-of-day tolling). Online strategies can be reactive or proactive (also known as anticipatory or predictive). Reactive strategies use observed traffic conditions to determine tolls, whereas proactive strategies combine both observed and predictive traffic conditions to determine the tolls.

Table 1. Recent Toll Optimization Literature

Study	Offline		Online	Analytical	Simulation-based	Toll	Network
	Static	Dynamic					
Lou et al. (2011)			✓	✓		1	One highway segment
Dong et al. (2011)			✓		✓	2	One part of the I-95 corridor and its surroundings with 3459 links
Hassan et al. (2013)			✓		✓	2	One corridor along with its adjacent arterials
Jang et al. (2014)			✓	✓		1	One 14-mile highway corridor in the San Francisco area
Toledo et al. (2015)			✓	✓		1	One 14-km highway corridor
Zheng et al. (2016)		✓		✓		1	Area-based toll for Sioux-Falls network
Chen et al. (2016)		✓			✓	5	Network of highways and arterials in Maryland (USA) with 2158 links
Gupta et al. (2020)			✓		✓	13	Network of Singapore's expressways and major arterials with 1150 links
Liu et al. (2017)	✓			✓		7	13-Link network
Han et al. (2017)		✓		✓		10	10-Link network
Vu et al. (2018)			✓		✓	16	Network of Singapore's expressways and major arterials with 1,150 links
Chen et al. (2018)		✓			✓	5	Network of highways and arterials in Maryland (USA) with 2,158 links
Zhang et al. (2018)			✓		✓	2	Highway corridor in Texas with 167 links
This paper	✓				✓	16	Network of Singapore's expressways and major arterials with 1,150 links

Table 1 summarizes some of the recent toll optimization literature. For each paper (i.e., each row), the table indicates whether it considers an offline static problem, an offline dynamic (i.e., time-dependent tolls) problem, or an online problem (whether reactive or proactive). The table also indicates whether the traffic model used for toll optimization is analytical or simulation based, the dimension of the toll vector, and a summary of the network of the largest case study in the paper.

The table indicates that the focus of recent work has mostly been on online (i.e., real-time) problems. Nonetheless, such approaches are mostly limited to simple applications that consider a single corridor with its neighboring arterials. Few approaches, including Gupta et al. (2020), Chen et al. (2016), and Vu et al. (2018), have considered more intricate and large-scale network topologies. Most literature has focused on low-dimensional problems. The recent works of Gupta et al. (2020) and Vu et al. (2018) consider a higher-dimensional problem that optimizes 13 and 16 tolls, respectively, distributed throughout Singapore.

The congestion pricing literature has extensively studied the importance of accounting for a detailed description of travel demand. For instance, accounting for the heterogeneity in the value of time across the population is important (Lou et al. 2011, Jang et al. 2014, Gupta et al. 2020). Furthermore, Vu et al. (2018) have enhanced the methods to account for elasticity of travel demand such that the travelers can change their mode and departure time and even cancel their trip in response to tolls. The work of Vu et al. (2018) considers distance-based tolling. The increased complexity of the demand modeling component has fostered the use of simulation-based models, which can embed detailed probabilistic travel demand models with random coefficients, such as value of time.

When the toll optimization problem is formulated as a simulation-based optimization problem, the most common approaches are the use of black box (i.e., general-purpose) algorithms such as genetic algorithms (Gupta et al. 2020, Vu et al. 2018) in combination with a constrained local search technique (Zhang et al. 2018). Such approaches can be directly used to address a variety of formulations (e.g., changes in the objective function or the feasible region can be readily accounted for). Nonetheless, this limits their performance under tight computational budgets or small samples. Given the computational cost of running high-resolution or large-scale traffic simulators, algorithms that can yield good quality solutions within few simulations are essential to address intricate optimization problems such as toll optimization.

The recent approach of Chen et al. (2016) has considered a metamodel approach to address the simulation-based optimization (SO) problem. A general-purpose Kriging metamodel is used to address a five-dimensional problem for a large-scale network with nonlinear network topology. They also apply similar simulation-based optimization techniques for improving the travel time reliability of the network (Chen et al. 2018). The advantage of using a general-purpose metamodel is that the approach can be directly applied to a variety of problem formulations.

Nonetheless, this generality also comes with the need to run a significant number of simulations prior to optimization. For instance, for the five-dimensional case study of Chen et al. (2016), a set of 100 points are simulated prior to optimization. This limits the use of the approach for high-dimensional problems. Distance-based toll optimization problems have also been receiving recent attention; Liu et al. (2017) provide a review. This paper tackles a problem with a decision vector of dimension 16. As indicated in Table 1, this is considered a high-dimensional problem in the toll optimization literature. However, it is not considered a high-dimensional problem for the broader field of continuous simulation-based optimization, where problems with approximately 100 variables are considered high-dimensional (Wang et al. 2016).

This paper focuses on the design of toll optimization problems with the following characteristics. First, we consider large-scale road networks such that the large-scale (e.g., across a full city or metropolitan region) impact of the tolls is accounted for. In particular, the use of a large-scale network model allows for capturing the impacts of the tolling on traffic assignment. This is particularly important when considering commuting patterns, where travelers with long-distance commutes will react to changes in tolls throughout the network. Second, we consider networks with intricate topologies. In particular, we aim to go beyond the analysis of linear-topology corridor studies. Third, we consider high-dimensional problems such that various tolls distributed throughout the network are simultaneously or jointly optimized. This is important such as to account for the global (or joint) impact of tolling on traffic assignment. This also allows for the coordination of tolls such as to account for equity considerations. Fourth, we focus on the use of stochastic, high-resolution, simulation-based traffic models. The latter allows for a detailed (e.g., probabilistic, dynamic) description of travel demand, which is essential to forecast the impact of tolls on congestion patterns.

This paper contributes to this area by focusing on the design of computationally efficient algorithms. These are algorithms that are designed to yield a good quality solution within a few simulation runs. The case study of this paper considers an offline static problem. We view the design of these computationally efficient offline algorithms as the building block for efficient real-time algorithms.

Computational efficiency can be improved through parallel computations such as in Gupta et al. (2020), Vu et al. (2018). Nonetheless, as our transportation systems and users become more real-time responsive and more connected, the intricacy of both the traffic simulation tools used and of the transportation optimization problems addressed increases. Hence, there is a need for computationally efficient algorithms.

Our approach to achieve computational efficiency is to allow the algorithms to exploit problem-specific structural information. More specifically, we propose to formulate an analytical network model that approximates the mapping between the tolls and the network-wide traffic conditions. We then embed this analytical structural information within the algorithm. In other words, this analytical information is combined with simulation-based information to identify suitable toll vectors.

This paper proposes a SO algorithm that enables high-dimensional toll optimization problems for large-scale networks to be addressed in a computationally efficient way. The essential component of the proposed methodology is the formulation of an analytical network model that provides an analytical and differentiable mapping between the toll vector and the network-wide performance metrics (such as revenue and traffic conditions). The proposed formulation is scalable and hence, suitable for large-scale networks. More specifically, for a network with n links, the analytical network model is formulated as a system of n nonlinear equations. Importantly, the analytical network model has endogenous traffic assignment, which is one of the main distinctions from our past work (Osorio and Bierlaire 2013, Osorio and Chong 2015). Nevertheless, its complexity (i.e., the dimension of the corresponding system of equations) scales independently of the dimension of the route choice set, of link attributes (e.g., link length, number of lanes), and of origin-destination matrix dimensions.

The proposed formulation is embedded within a metamodel SO algorithm and is used to address a high-dimensional offline toll optimization problem for a large-scale Singapore network. More specifically, 16 tolls that are distributed throughout the network of Singapore expressways and major arterials are optimized. The network is modeled as set of more than 1,150 links, 2,300 lanes, and 4,050 OD pairs with more than 18,000 routes.

Section 2 presents the proposed methodology. Validation experiments are presented in Section 3, followed by a Singapore case study in Section 4. Conclusions are discussed in Section 5.

2. Methodology

2.1. Problem Formulation

To formulate the toll optimization problem, we introduce the following notation:

$\boldsymbol{x}$	decision vector (i.e., toll vector) in a given currency unit;
$f(\boldsymbol{x})$	simulation – based objective function;
$F_i(\boldsymbol{x})$	hourly flow on link i;
$E[F_i(\boldsymbol{x})]$	expected hourly flow on link i;
$\boldsymbol{x}_L$	lower bound vector;
$\boldsymbol{x}_U$	upper bound vector;
$\mathcal{T}$	set of links with tolls;
$\mathcal{I}$	set of all links.

The problem is formulated as follows:

$$\max_{x} f(\boldsymbol{x}) = \sum_{i \in \mathcal{I}} x_i E\left[F_i(\boldsymbol{x})\right] \quad (1)$$

$$\boldsymbol{x}_L \leq \boldsymbol{x} \leq \boldsymbol{x}_U. \quad (2)$$

The decision vector $\boldsymbol{x}$ is the vector of toll rates, with element i denoted by x_i. For the case study of Section 4, the dimension of $\boldsymbol{x}$ is 16. In other words, the tolls of 16 distinct locations are determined simultaneously. The objective function $f(\boldsymbol{x})$ (Equation (1)) is an unknown simulation-based function, which in this paper is the expected hourly revenue. Term i in the summation represents the expected hourly revenue for toll i. The latter is defined as the product of the toll and of the expected number of vehicles per hour that travel on link i. This expected link flow is a function of the tolls because the tolls impact traffic assignment (i.e., when choosing their routes, the travelers account for the toll costs). This expectation is an unknown function, which is estimated via simulation. The toll optimization problem considers lower and upper bound constraints for the tolls (Equation (2)). These bounds are available in analytical form; that is, they are not simulation-based constraints. This problem formulation considers the main incentive from the perspective of the toll operator: revenue maximization. The proposed methodology is suited to address other problem formulations that account explicitly for how tolls impact network performance. This can be done by including in the objective function additional network-wide performance metrics such as travel times and speeds. In summary, the problem is a SO problem with continuous decision variables, simulation-based objective function, and analytical deterministic bound constraints. Other problem formulations that account, for instance, for the level of service or the social welfare can also be addressed with the proposed methodology.

In practice, dynamic tolls throughout a network are determined independently (i.e., the spatial dependencies between tolls is not explicitly accounted for in the optimization). Commonly used approaches in practice are rule-based approaches or lookup tables that determine tolls based on the prevailing traffic conditions. An example is the dynamic tolling on I-394 in the Minneapolis-St. Paul, MN, metropolitan area (Halvorson et al. 2011). Recent literature has addressed the spatial dependencies between tolls. However, many recent studies focus on the analysis of a single toll, as illustrated in Table 1. The proposed formulation simultaneously determines all tolls in the network. In the Singapore case study of this paper, a set of 16 tolls distributed throughout Singapore are jointly or simultaneously determined. This allows us to account for the joint impact of tolls on the spatial and temporal propagation of congestion. For instance, a toll on an upstream link influences flow on downstream links, which influences downstream tolls. This illustrates the importance of accounting for these dependencies between the tolls for toll optimization. More generally, it illustrates the intricate mapping between the toll vector, the underlying traffic dynamics, and the resulting objective function.

In practice, the toll bounds (Equation (2)) are typically given by the context. The upper bound is usually set by regulations. It is not a decision variable. As an example of the discussion of such regulations, we refer to Albalate et al. (2009), and an example from the U.S. context is given by Zhang et al. (2018). The lower bound can be set based on tolling policies. For example, it can be set such as to avoid large toll fluctuations across time intervals.

This paper considers static tolls. In other words, for a given time period (e.g., weekdays, morning peak period) the toll costs are determined offline and fixed. The traffic simulator used to determine the tolls is a dynamic traffic model. The use of a dynamic simulation model, unlike a static one, enables a more detailed description of vehicle-to-vehicle interactions. Therefore, it can provide a more detailed description of the spatial and temporal propagation of congestion and a more accurate estimation of travel times. Arguably, the most important factors when determining tolls are the value of time of users and the travel time estimates. The simulation model used accounts for heterogenous value of time of the traveler population. Hence, as congestion levels and travel times vary across time, the proportion of users that are willing to pay the toll may also vary. A dynamic model combined with a route choice model that embeds heterogenous value of time allows us to capture this.

2.2. Metamodel SO Algorithm

We use in this paper the general metamodel SO approach of Osorio and Bierlaire (2013). To briefly describe its main ideas, we introduce the following notation. The index k refers to a given SO algorithm iteration.

m_k	metamodel function;
$\boldsymbol{\beta}_k$	parameter vector of metamodel m_k;
$\beta_{k,j}$	element j of the parameter vector $\boldsymbol{\beta}_k$;
$f_A(\boldsymbol{x})$	approximation of the SO objective function provided by an analytical traffic model;
$\phi(\boldsymbol{x};\boldsymbol{\beta}_k)$	polynomial component of the metamodel m_k;
$h(\boldsymbol{x};p)$	analytical traffic model formulated as a system of nonlinear equations.

To address Problem (1)-(2), each iteration k of the SO algorithm solves a metamodel optimization problem of the following form:

$$\max_{x} m_k(\boldsymbol{x};\boldsymbol{\beta}_k) = \beta_{k,0} f_A(\boldsymbol{x}) + \phi(\boldsymbol{x};\boldsymbol{\beta}_k) \quad (3)$$

$$h(\boldsymbol{x};p) = 0 \quad (4)$$

$$\boldsymbol{x}_L \leq \boldsymbol{x} \leq \boldsymbol{x}_U. \quad (5)$$

Problem (1)-(2) differs from Problem (3)-(5) in two main ways. First, the SO objective function, $f(\boldsymbol{x})$, is replaced by

an analytical and differentiable function known as the metamodel m_k. The latter is a parametric function that is defined as the sum of: (i) an analytical approximation of $f(\boldsymbol{x})$ provided by an analytical traffic model ($f_A(\boldsymbol{x})$ term) and (ii) a polynomial function (ϕ term). The metamodel can be seen as the approximation of f provided by an analytical traffic model and corrected for with both a scaling factor ($\beta_{k,0}$) and an additive error term (ϕ term). The polynomial, ϕ, is defined as a quadratic function with diagonal second-derivative matrix. More specifically,

$$\phi(\boldsymbol{x};\boldsymbol{\beta}_k) = \beta_{k,1} + \sum_{j=1}^{T} x_j \beta_{k,j+1} + \sum_{j=1}^{T} x_j^2 \beta_{k,j+T+1}, \qquad (6)$$

where T is the dimension of the decision vector $\boldsymbol{x}$. Equation (6) defines the function ϕ of (3), which is the general-purpose (i.e., not problem-specific) component of the metamodel. Second, it has an additional set of constraints (Equation (4)) that represent the analytical traffic model. The latter is formulated as a system of nonlinear equations.

Each iteration k of the SO algorithm carries out the following main steps: (i) use all simulation observations collected so far (i.e., all estimates of $f(\boldsymbol{x})$) to fit the parameter vector, $\boldsymbol{\beta}_k$, of the metamodel (the least-squares problem that is solved to fit $\boldsymbol{\beta}_k$ is detailed in the online appendix; it aims to minimize a distance function between the metamodel predictions and the simulated estimates of $f(\boldsymbol{x})$); (ii) solve the metamodel optimization problem (3)-(5); and (iii) simulate new points (for instance, simulate the optimal solution of the metamodel optimization problem).

The main component of the metamodel that can enable it to address high-dimensional and intricate SO problems (e.g., nonconvex) in a computationally efficient way is the analytical traffic model approximation (i.e., f_A term). This term provides a problem-specific approximation of f, whereas the polynomial provides a general-purpose approximation. In other words, depending on the choice of f (e.g., revenue, consumer surplus, etc.), the functional form of f_A will vary, whereas that of ϕ will not. The analytical traffic model provides analytical structural information to the SO algorithm. More specifically, it provides an analytical and physically plausible (i.e., problem-specific) approximation of the mapping between the decision vector and the objective function. Traditional SO algorithms treat the simulator as a black box. The use of f_A enables problem- and network-specific information to be provided to the algorithm. Hence, the simulator is no longer treated as a black box.

The main challenge in this metamodel SO approach is the formulation of an analytical traffic model that has the following properties. First, it should provide a good approximation of the unknown simulation-based objective function f (such that the optimal solutions to Problem (3)-(5) are good quality solutions to Problem (1)-(2)). Second, it should provide a good global approximation (i.e., a good approximation in the entire feasible region). This differs from local metamodels such as polynomials. Third, it should be computationally efficient to evaluate. Because Problem (3)-(5) is solved at every iteration of the SO algorithm, it needs to be solved efficiently (otherwise, we are better off allocating the computing resources to running additional simulations). Fourth, it should be a scalable model such that it can be used for large-scale networks, that is, the system of equations (4) needs to be efficiently evaluated for large-scale networks. Fifth, it should be differentiable such that Problem (3)-(5) can be solved with a variety of traditional gradient-based algorithms. In Section 2.3, we formulate an analytical traffic model with all of the above properties for toll optimization problems. This general metamodel SO idea has been formulated and used to design efficient algorithms for various transportation problems, including various traffic signal control problems (Osorio and Chong 2015, Osorio and Nanduri 2015, Chong and Osorio 2018, Chen et al. 2019), and more recently for model calibration problems (Zhang et al. 2017).

2.3. Traffic Model Formulation

To formulate the analytical traffic model, we introduce the following additional notation:

Endogenous variables of the analytical traffic model :

y_i	expected hourly demand per lane of link i;
k_i	expected density per lane of link i;
v_i	expected$(space-mean)$speed per lane of link i;
t_r	expected travel time for router;
z_r	toll cost for router;
$P(r)$	route choice probability for router.

Exogenous parameters of the analytical traffic model :

d_s	expected hourly travel demand for OD pairs;
k_i^{jam}	jam density per lane of link i;
$v_i^{\max}$	maximum speed of link i;
q^{cap}	lane flow capacity;
n_i	number of lanes of link i;
ℓ_i	average lane length of link i;
θ_1, θ_2	coefficients of the route choice model;
$\alpha_{1,i}, \alpha_{2,i}$	parameters of the fundamental diagram of link i;
c	scaling parameter common to all links;
$O(r)$	OD pair of router;
$\mathcal{R}_1(i)$	set of routes that include link i;
$\mathcal{R}_2(s)$	set of routes of OD pairs;
$\mathcal{R}$	set of all routes;
$\mathcal{L}(r)$	set of links of router;
$\mathcal{T}$	set of links with tolls;
$\mathcal{I}$	set of all links.

The analytical traffic model is formulated as follows:

$$\begin{cases} f_A(\boldsymbol{x}) = \sum_{i\in\mathcal{I}} x_i n_i y_i & \text{(7a)} \\ y_i = \frac{1}{n_i} \sum_{r\in\mathcal{R}_1(i)} P(r) d_{O(r)} \quad \forall i \in \mathcal{I} & \text{(7b)} \\ P(r) = \frac{e^{\theta_1 t_r + \theta_2 z_r}}{\sum_{j\in\mathcal{R}_2(O(r))} e^{\theta_1 t_j + \theta_2 z_j}} \quad \forall r \in \mathcal{R} & \text{(7c)} \\ t_r = \sum_{i\in\mathcal{L}(r)} t_i \quad \forall r \in \mathcal{R} & \text{(7d)} \\ z_r = \sum_{i\in\mathcal{L}(r)\cap\mathcal{T}} x_i \quad \forall r \in \mathcal{R} & \text{(7e)} \\ t_i = \frac{l_i}{v_i} \quad \forall i \in \mathcal{I} & \text{(7f)} \\ v_i = v_i^{\max}\left(1 - \left(\frac{k_i}{k_i^{\text{jam}}}\right)^{\alpha_{1,i}}\right)^{\alpha_{2,i}} \quad \forall i \in \mathcal{I} & \text{(7g)} \\ k_i = c\frac{k_i^{\text{jam}}}{q^{\text{cap}}} y_i \quad \forall i \in \mathcal{I} & \text{(7h)} \end{cases}$$

Equation (7a) is the approximation of the expected hourly revenue provided by the analytical traffic model. It is the analytical counterpart of Equation (1), where the simulation-based expected hourly flow of link i, $E[F_i(\boldsymbol{x})]$, is replaced by the analytical expected hourly demand of link i. The latter is defined as the product of the hourly demand per lane of link i, y_i, and the number of lanes on link i, n_i. The hourly demand per lane of link i is defined by Equation (7b) as the ratio of the sum of the expected demand for all routes that travel through link i and of the number of lanes of link i. The expected route demand is the product of the route choice probability and the total demand for the underlying OD (Origin-Destination) pair. The route choice probability (Equation (7c)) is approximated as a multinomial logit model with a utility that is a function of the route's expected travel time and the route's toll cost. The expected route travel time (Equation (7d)) is the sum of the expected travel times of its links. The route toll cost is the sum of the tolls that are on the given route (Equation (7e)). Equation (7f) defines the expected link travel time as the ratio of the link length and the expected (space-mean) link speed. The fundamental diagram of the link is given by Equation (7g). It relates link densities to link speeds. Finally, Equation (7h) relates the link demand to the link density. The underlying assumption is that the ratio of density to jam density is proportional to the ratio of hourly demand to flow capacity. The proportionality constant, c, is common for all links. In this model, a common flow capacity value is used for all links. This system of equations (7) defines the system of equations denoted h in Equation (4).

In summary, the above analytical traffic model provides a simplified (compared with the simulator) description of how tolls impact the spatial distribution of vehicular flow. The system of equations (7) accounts for the dependencies between tolls (recall the discussion in Section 2.1) through an analytical description of how tolls impact route choices, which in turn impacts the spatial propagation of congestion and the corresponding toll revenues.

The above model (Equation (7)) is implemented for a network with n links as a system of n nonlinear equations. In other words, the complexity of the model (i.e., the dimension of the corresponding system of equations) scales linearly with the number of links in the network and is independent of other link attributes such as link length or lane attributes. What is particularly remarkable is that the analytical model has endogenous traffic assignment (i.e., endogenous route choice), yet its complexity does not depend on the size of the route choice set. For instance, for the Singapore case study of this paper, the analytical model considers a network with 860 links, 4,050 OD pairs, and more than 18,200 routes and is implemented as a system of 860 nonlinear equations. This is a scalable formulation suitable for large-scale network optimization.

A description of how this analytical model differs from the simulation model used for the case studies of this paper is given in the online appendix. The exogenous parameters of the analytical model are calibrated as follows.

The simulator considers multilane links, with heterogeneous lanes, whereas the analytical model assumes all lanes of a link to be homogeneous. Hence, the link attributes ($k_i^{\text{jam}}, v_i^{\max}$ and ℓ_i) of the analytical model are estimated as the average, over all lanes of the link, of the individual lane attributes of the simulator. The parameters of its fundamental diagram of each link of the analytical model are defined based on those of the simulator; the specific equation that governs this is detailed in the online appendix. For all lanes, the analytical model uses a single lane flow capacity value (denoted q^{cap}), and its value is set as the maximum flow capacity value, among all of the simulators lanes. The simulator has a predetermined route choice set, which defines the route choice set of the analytical model. Additional details on this are provided in the online appendix. The simulator allows for time-dependent OD demand matrices. The time-independent OD demand, d, of the analytical model is obtained as the average, over the time period of interest, OD demand. For the case studies of this paper, the proportionality coefficient c of Equation (7h) is kept to the same value, which is set to $1/6$. This value was obtained through insights from numerical experiments on a toy network. As is described in the online appendix. the simulator uses a route choice model with a probabilistic value of time, whereas the analytical model considers a deterministic value of time. The coefficients of the route choice model (denoted θ_1 and θ_2), including the value of time, are estimated as the

expected value of the probabilistic coefficients of the simulator's route choice model.

3. Validation

We carry out experiments on a synthetic small toy network, as illustrated in Figure 1. The network has one OD pair and three multilane links in a diverge node topology. In other words, all trips have a common origin link. After traveling on link 1, the travelers can choose between links 2 and 3. The only difference between links 2 and 3 is that link 2 is tolled. This is a one-dimensional problem (i.e., a scalar toll value). We consider experiments with two different value of time (VOT) values (\$15/h and \$30/h). For each of the experiments, we consider four different demand scenarios. Three demand scenarios have a constant OD demand with values of 3,600, 4,800, and 6,000 vehicles per hour, respectively. The fourth demand scenario considers time-dependent OD matrices, where demand increases gradually (from 3,600 to 6,000), and the average demand is 4,800 vehicles per hour. These experiments cover various levels of congestion ranging from free-flowing conditions to congested conditions.

Figure 2 displays eight plots. The top plots display the objective function (i.e., expected revenue function f of Equation (1)) estimates obtained via simulation. The bottom plots display the approximations derived by the analytical network model (i.e., function f_A of Equation (3)). A given row of plots displays, from left to right, the demand scenarios with constant demand of 3,600, constant demand of 4,800, constant demand of 6,000, and increasing demand with an average of 4,800, respectively. Each plot displays two curves that correspond to a given VOT value; a VOT of 15 corresponds to the solid curve, and a VOT of 30 corresponds to the dashed curve. Note that the different plots have different y-axis limits. For a given demand scenario (i.e., a given plot), as the VOT increases, the maximum revenue increases, and the toll for which maximum revenue is obtained also increases. This holds for all demand scenarios. This trend is well replicated by the analytical model. For a given demand scenario (i.e., when comparing a given column of plots), the analytical model has an accurate approximation of the value of the optimal toll. This holds for all demand scenarios, with both constant and time-dependent demand. Nonetheless, the magnitude of the revenue functions differ. The analytical model tends to overestimate the simulated revenue.

Figure 3 considers the same set of experiments. It now displays, in the same plot, the experiments with common VOT value yet different demand scenarios. As before, the top (resp. bottom) plots correspond to simulation-based estimates (resp. analytical approximations). The four demand scenarios are represented as follows: constant demand of 3,600 (blue curve), constant demand of 4,800 (black curve), constant demand of 6,000 (green curve), and increasing demand with an average of 4,800 (red curve). The top plots indicate that for a given VOT value, as the average demand increases, so does the maximum revenue, yet the value of the optimal toll does not vary much. The analytical model captures these trends, as indicated by the bottom plots. Because the analytical model is time invariant (in particular, it is a stationary model), the revenue values for both demand scenarios (constant demand and increasing demand) with an average demand of 4,800 are identical. The simulator, which is a dynamic model, also yields similar expected revenue functions for these two demand scenarios.

We address the toll optimization problem for this toy network. We consider a VOT value of \$15/h and a constant demand of 4,800 vehicles per hour. We compare our proposed approach with the derivative-free generalized pattern search algorithm (Mathworks, Inc. 2016). This algorithm was chosen as a benchmark for the following reasons. As a derivative-free algorithm, it does not rely on first-order derivative information, which can be computationally costly to estimate. It is suitable for noncontinuous and nondifferentiable objective functions, as is the case of our simulation-based objective function (the traffic simulator relies on noncontinuous and nondifferentiable functions). Moreover, recent work has highlighted the good performance of similar direct search

Figure 1. Toy Network Topology

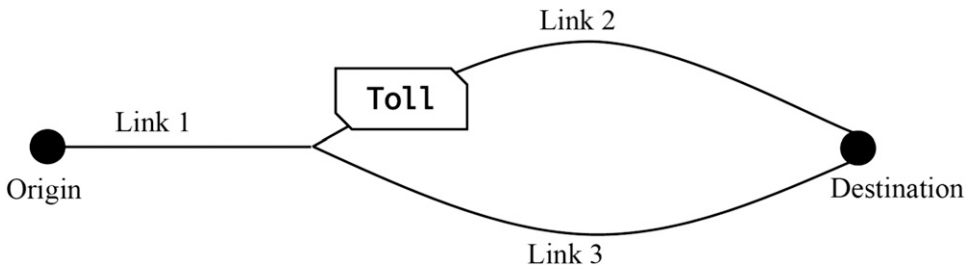
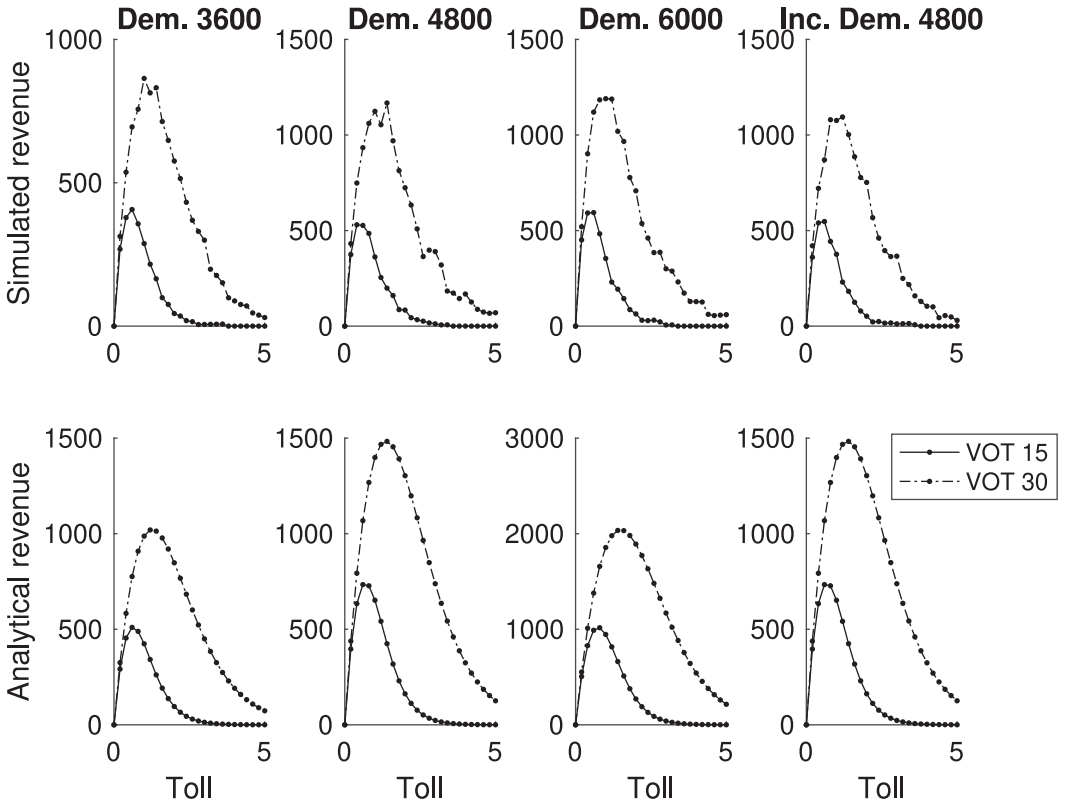


Figure 2. Comparison of the Simulation-Based Objective Function and the Analytical Objective Function for Various Demand and Value of Time Scenarios



techniques when used under tight computational budgets (Dong et al. 2017).

We consider 20 different initial points, which are uniformly drawn from the feasible region (Equation (2)). For each initial point, we run each method (i.e., algorithm), allowing for a computational budget of 50 simulations (i.e., the algorithm is terminated once a total of 50 simulations are evaluated).

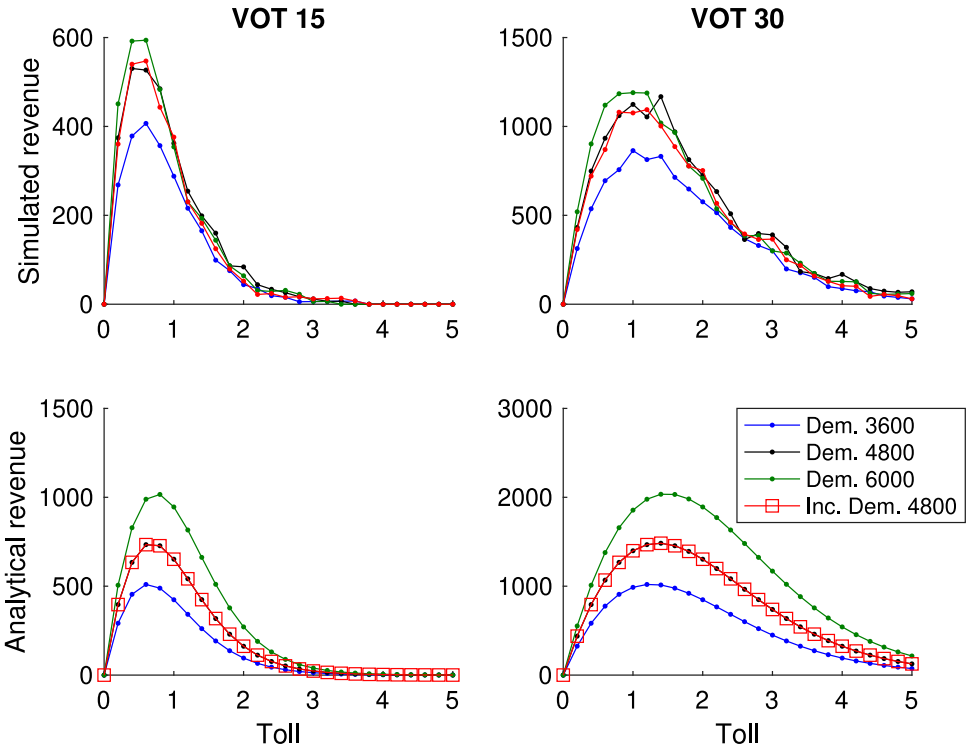
Figure 4 displays the performance of each method considering each of its 20 runs. Each plot displays the estimate of the objective function of the current iterate (i.e., the point considered to have best performance so far) as a function of the total number of simulated points (i.e., computational budget used so far). The top (resp. bottom) plot considers the proposed (resp. benchmark) method.

The top plot indicates that all 20 runs of the proposed method identify at iteration 1 a solution with excellent performance. This shows the ability of the proposed method to perform well regardless of the quality of the initial solution. In other words, it is robust to the quality of the initial point. The benchmark method gradually finds points that have improved performance compared with the initial point. Its performance for low computational budgets (e.g., 10 simulation runs) is sensitive to the quality of the initial point. For larger budgets, it has a performance similar to that of the proposed method. All 20 runs of the proposed method converge to the same final solution, which is the global optimal solution. For the benchmark method, 16 of the 20 runs converge to this value, and the other four runs converge to a local optimum with an estimated objective function that is 4% lower than the global optimal solution.

4. Singapore Case Study

We consider a case study with a network model of the major arterials and expressways of Singapore as seen in Figure 5(a). We consider a set of 16 tolls to optimize. The 16 toll gantries are distributed throughout the network and labeled in the figure with red circles. The expressway links are represented with orange lines, and the major arterials are represented with yellow lines. The simulation model of the network is displayed in Figure 5(b). It accounts for all expressway links and a few arterial links. Currently, Electronic Road Pricing (ERP) is deployed in Singapore such that travelers are charged at gantries when they enter a zone. The toll to be charged varies across the day (time-of-day tolling) and location. The pricing is

Figure 3. (Color online) Comparison of the Simulation-Based Objective Function and the Analytical Objective Function for Various Demand and Value of Time Scenarios



predetermined; it is reviewed quarterly and published so that the travelers know exactly the toll they will pay. We refer to Seik (2000) for the history of the ERP system in Singapore.

The network model consists of 1,150 links with more than 2,300 lanes, 4,050 OD pairs with more than 18,000 routes. Note that of the 1,150 links of the simulator, only 860 belong to routes of the predetermined route choice set. Details on how this choice set is defined are given in the online appendix.

Hence, the analytical model accounts only for those 860 links. Demand is defined, in the simulator, as calibrated, time-dependent OD matrices for every five-minute interval. We simulate a weekday 8 to 9:30 a.m. period; during this period, expected travel demand is of the order of 226,247 trips. We determine the tolls such as to optimize revenue for the 9 to 9:30 a.m. period. We set the lower toll bound to its smallest value of S$0 and upper toll bound to a relatively high value of S$5 (Singapore dollars). As a reference, toll values in Singapore for the period of May to August 2018 are provided by the Land Transport Authority (2019), and toll values tend to be significantly lower than S$5. We consider five initial points. Each initial point is uniformly drawn from the feasible region (Equation (2)). For a given point, we run each algorithm once and allow for a computational budget of 80 simulations.

We compare the performance of the algorithms as a function of the number of simulated points (i.e., the amount of computational budget depleted). This allows for a comparison that is hardware (e.g., computational resources) and software (e.g., code implementation) agnostic. We also compare their performance as a function of the total computation time. The runs

Figure 4. (Color online) Comparison of the Performance of the Algorithms for a Toy Network Toll Optimization Problem

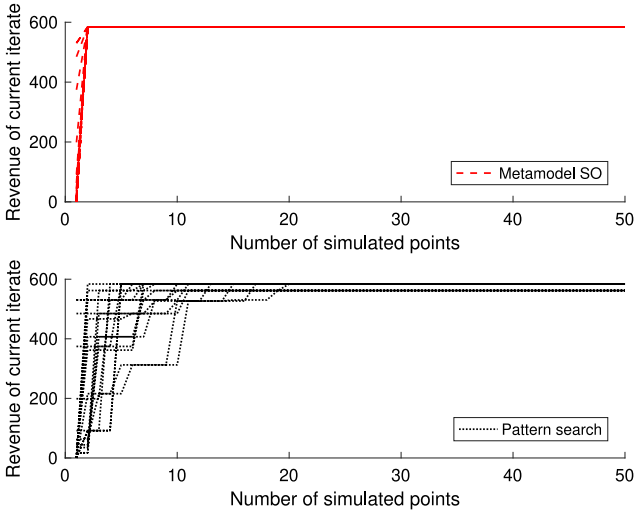
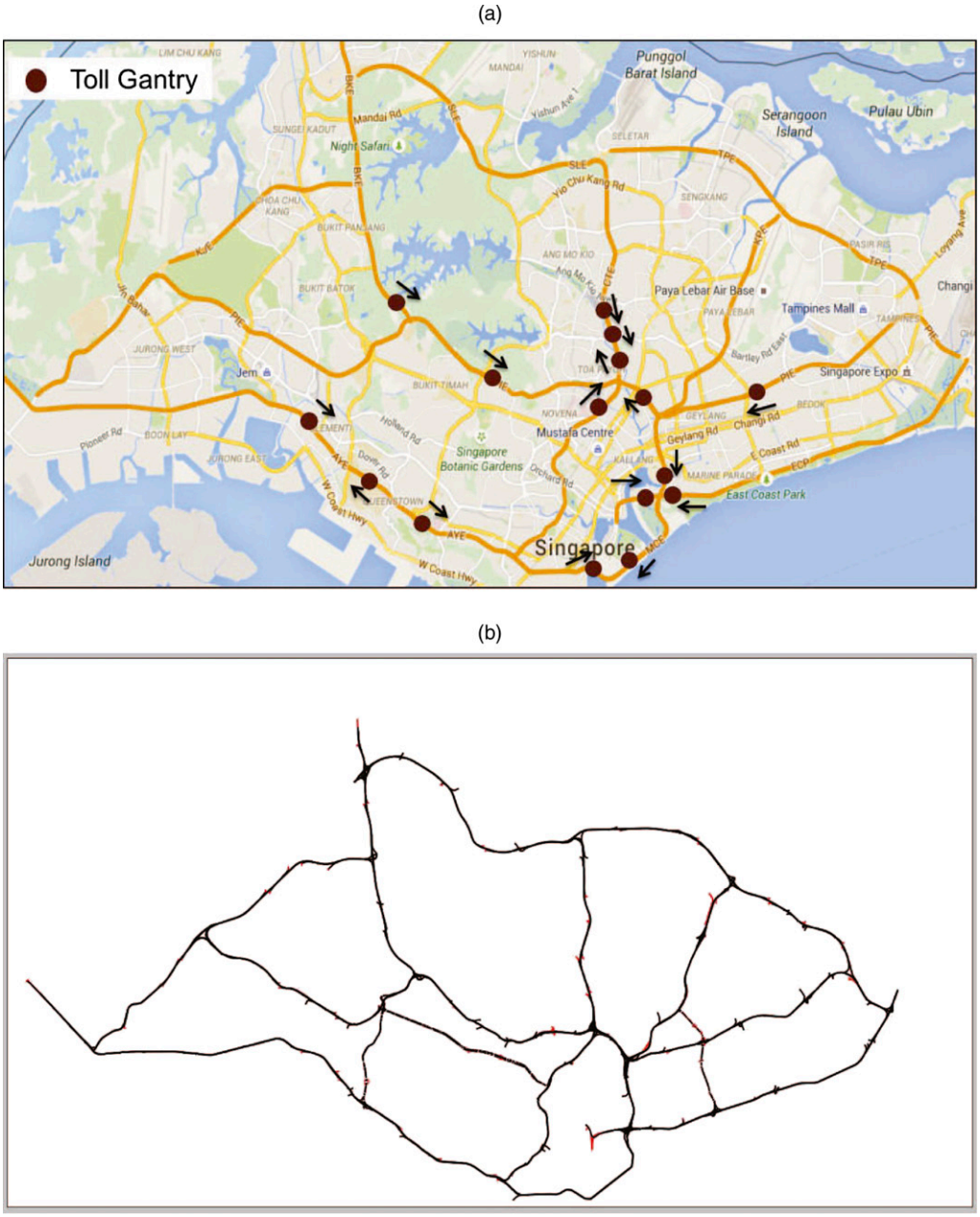


Figure 5. (Color online) Singapore Network

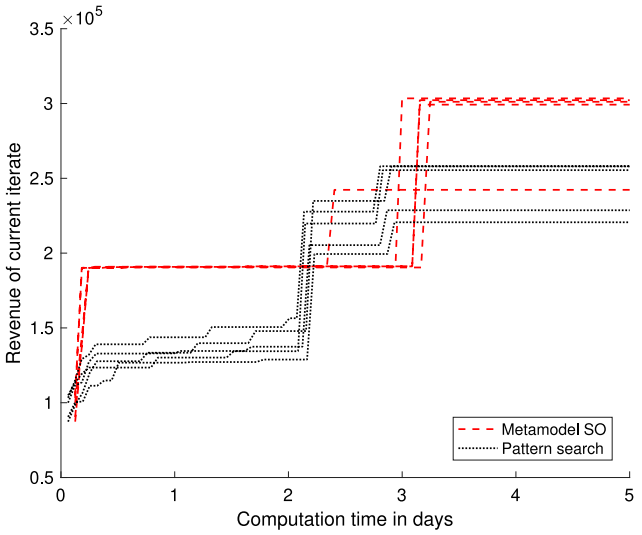


Notes. (a) Singapore expressway network (map data: Google Maps 2017). (b) Simulation network model.

are carried out on various servers, which do not guarantee equal CPU allocation to each job. Hence, the computation times are indicative of efficiency, yet the differences may not be solely attributable to the algorithm.

Figure 6 displays the performance of each algorithmic run as a function of computation time. The red dashed lines represent the proposed metamodel SO method. The black dotted lines represent the benchmark method. For each algorithm, there are five lines, which correspond to five different initial points. The x-axis displays the total computation time in days. The y-axis displays the simulation-based estimate of the objective function of the current iterate. Note that on average one simulation run takes approximately two hours to compute. Figure 6 indicates that all five runs of the proposed method significantly outperform the benchmark runs for the first two days of computation. Then, there is a phase where all runs of the benchmark outperform four runs of the proposed

Figure 6. (Color online) Revenue of the Current Iterate as a Function of the Cumulative Computation Time



method. After day 3.5, four of the proposed method runs outperform all benchmark runs. The fifth run outperforms two of the five benchmark runs.

This figure illustrates that the proposed method often (in four out of five instances) identifies final solutions that significantly outperform the solutions proposed by the benchmark method. When comparing the objective function value of the final solution and averaging over the five runs, the proposed method yields an average improvement of 19% (average revenues of 289,678 vs. 244,158). Similarly, when comparing the revenue of the best solution (out of the 5 solutions) proposed by each method, the proposed method yields an improvement of 18% (revenue of 303,450 vs. 258,090). The best solution is defined as that with the highest objective function estimate. Four of the five proposed solutions (i.e., solutions derived by the proposed method) have very similar objective function values. This indicates robustness of the method to both the quality of the initial point and the stochasticity of the simulator. There is higher variability among solutions derived by the benchmark method.

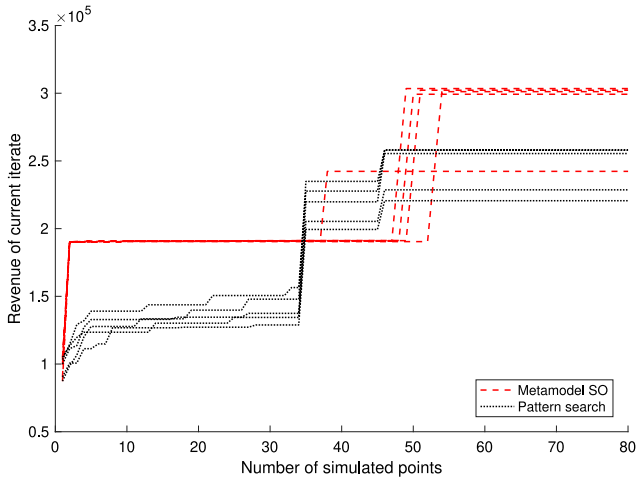
During the first two days of computation, the benchmark method slowly finds solutions that gradually improve performance. As of day 2, it finds solutions with performance comparable to those of the proposed method and that significantly outperform the initial solutions. On the other hand, the proposed method immediately (i.e., at iteration 1) identifies solutions with significantly improved performance compared with the initial solution (at iteration 1, the revenue improves on average by 95%; i.e., average revenues of 190,150 vs. 97,229). This initial improvement is entirely due to the analytical network model. More specifically, when the algorithm starts, there are no simulation observations available; hence, the first current iterate is defined as the solution to the analytical network model problem (i.e., a problem that maximizes $f_A(\boldsymbol{x})$ of Equation (3) subject to constraints (4)-(5)). This shows the added value of the structural information provided by the analytical network model. Also note that the solutions to this analytical network model problem are the same for all five initial points. This shows the added value of using an analytical network model that provides a good global approximation of the objective function (rather than using local models, such as polynomials). It is this analytical network model that leads to a SO algorithm that is robust to the quality of the initial points.

Figure 7 differs from the previous figure in that it considers the performance as a function of the total number of simulated points (instead of the total computation time) (i.e., the two figures differ in their x-axis). This figure serves to compare the performance of the algorithm independently of hardware and software considerations. The lines are almost identical to those of the previous figure. All conclusions from the previous figure also hold here.

Figure 8 considers for each algorithm the five solutions it yields upon termination and displays the coordinates of the solutions (i.e., toll values). Each plot contains 16 boxplots, one for each toll (i.e., each i value). For a given i, the boxplot illustrates the variability (across the 5 solutions) of the toll values, x_i. The top, middle, and lower plot display, respectively, the five initial points, the five final solutions of the proposed method, and the five final solutions of the benchmark method. The top plot confirms that the five initial points are very different (i.e., high variability). The middle plot shows that for most tolls there is little variability across toll values. More specifically, 13 of the 16 tolls have low variability (the tolls with high variability are indexed 5, 8, and 16). The lower plot shows that the benchmark method yields solutions with lower variability than the initial point, but with higher variability than the proposed method. More specifically, 10 of the 16 tolls have lower variability under the proposed method than under the benchmark method. For the proposed method, eight of the tolls have a variance smaller than 0.1, whereas for the benchmark method, only one of the tolls is below this threshold.

Figure 9 considers all of the points simulated across all 10 runs (5 for the benchmark method and 5 for the proposed method). For each of these points, it evaluates the analytical approximation of the objective function provided by the analytical network model. In other words, it considers a given $\boldsymbol{x}$ and solves the system of nonlinear Equations (7) to obtain $f_A(\boldsymbol{x})$, which is the analytical network model approximation of $f(\boldsymbol{x})$

Figure 7. (Color online) Revenue of the Current Iterate as a Function of the Number of Simulated Points



of Equation (1). For each simulated point, x, Figure 9 displays $f_A(x)$ along the x-axis and the simulated estimated of the expected revenue (i.e., an estimate of $f(x)$) along the y-axis. Each point (i.e., each x value) is displayed as a cross. The figure also displays the diagonal line (dashed blue line). This figure shows that most points lie along a line with positive slop. This indicates that the analytical network model provides approximations that have high positive correlation with the simulation-based estimate. It is remarkable that such a simple analytical formulation captures the main trends of such an intricate simulation-based objective function.

Figure 8. (Color online) Variability of the Toll Vectors for the Initial Points and the Solutions Proposed by Each Method

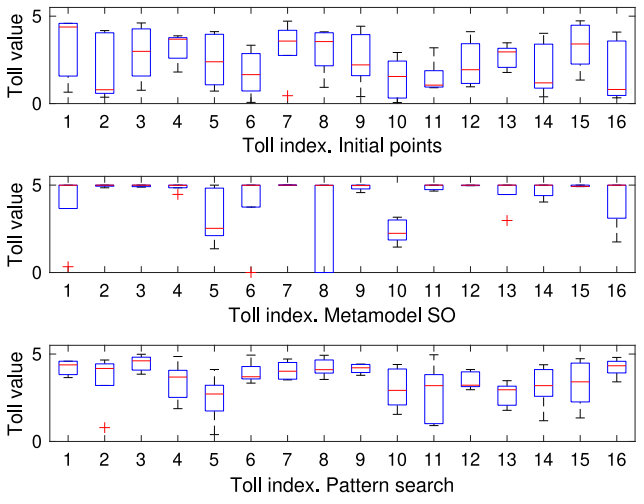
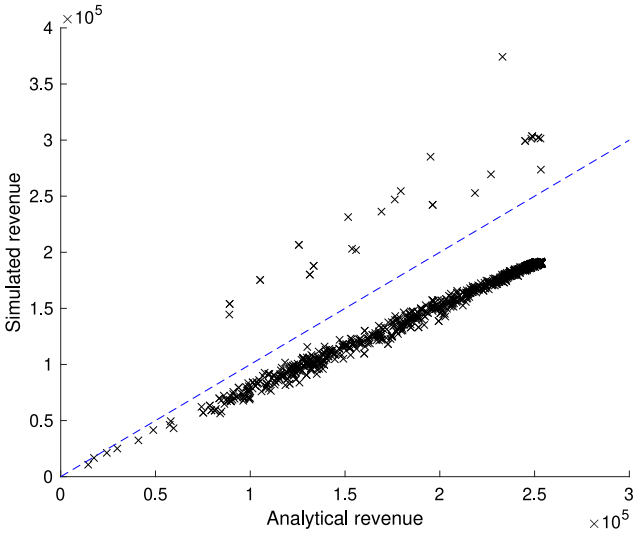


Figure 9. (Color online) Comparison of the Simulated Revenue and the Revenue Approximated by the Analytical Network Model



5. Conclusions

This paper proposed a simulation-based optimization technique for high-dimensional toll optimization problems of large-scale road networks. The main novelty is the formulation of an analytical network model. The latter is embedded within a metamodel simulation-based optimization (SO) algorithm. It provides analytical and differentiable structural information of the underlying problem to the SO algorithm. Hence, the algorithm no longer treats the simulator as a black box.

The analytical model is formulated as a system of nonlinear equations that can be efficiently evaluated with standard solvers. The dimension of the system of equations scales linearly with network size and independently of the dimension of the route choice set and of link attributes, such as link length. Hence, it is a scalable formulation suitable for the optimization of large-scale networks. For instance, the model is used in Section 4 for toll optimization of a Singapore network with more than 4,050 OD pairs and 18,200 feasible routes. The corresponding analytical model is implemented as a system of 860 nonlinear equations.

The analytical network model is validated based on one-dimensional problems of a toy network. It captures the main trends of the simulation-based objective function and, more importantly, accurately locates the global optimum for all experiments. The proposed SO approach is then used to optimize a set of 16 tolls for the network of expressways and major arterials of Singapore. The proposed approach identifies good quality solutions at the very first iteration of the algorithm. This is entirely due to the analytical structural information provided by the analytical network model. More specifically, at iteration 1, the

proposed method improves the objective function value by 95%. The benchmark method is slow to identify solutions with good performance; this is achieved after 2 days of computation or similarly after 30 points are simulated. The proposed algorithm yields final solutions that outperform those of the benchmark method by an average of 18%. The experiments also indicate that the analytical structural information enables the SO algorithm to become robust to both the quality of the initial points and the simulator's stochasticity.

This general idea of formulating and using structural analytical problem-specific information is promising for the design of real-time algorithms. Cities worldwide are increasingly collecting real-time data and adapting their operations to respond to the real-time variations of demand and of supply. Hence, there is a pressing need to develop algorithms that both are real-time feasible and capture the network performance at a large-scale (e.g., an entire city or metropolitan region). Coupling general-purpose black box techniques (e.g., machine learning, simulation-based optimization) with such analytical structural information is a promising approach to enable them to achieve real-time feasibility for both high-dimensional problems and large-scale networks. Such structural information can also improve the robustness of these methods (e.g., to measurement errors, to sparse data).

There are various ways in which the structural information of the analytical traffic model can be used to enhance the performance of real-time algorithms. One promising approach is to use the analytical approximation of the simulation-based objective function to design a sampling distribution that balances the traditional exploration-exploitation trade-off. Ideas along these lines have been developed, for simulation-based signal control problems, in Tay and Osorio (2017). This allows us to simulate both (i) points with good performance (i.e., good objective function estimates) and (ii) points that further explore the feasible region.

The majority of the computational effort lies in the evaluation of the simulation model. For online optimization, it is essential to find more efficient ways of performing simulation evaluations. The traditional approach is to focus on parallelization. Promising, and recently proposed ideas, include marginal simulation (Corthout et al. 2014) and multimodel optimization (Osorio and Selvam 2017).

This paper focuses on the design of algorithms that can identify points with good performance within a tight computational budget (i.e., within few simulation runs). Hence, the algorithm is terminated once this computational budget is reached (i.e., once a maximum number of simulations are performed). This does not guarantee a locally optimal solution. If larger computational budgets are considered (e.g., at least 2 orders of magnitude higher), then the algorithm can incorporate local optimality stopping criteria. Performing such local optimality tests is computationally demanding. Therefore, it is essential to devise strategies that trade off the frequency of performing these tests with the associated computational costs. A discussion of these trade-offs in the context of discrete SO is discussed in Xu et al. (2013).

In practice, changes in tolls can impact not only route choices but also departure time and mode choices. The simulator used in this paper can account for changes of departure time, mode, and trip cancellations. This can also be accounted for in the analytical model, just as we did in this paper for route choice. This entails using a simplified mode and/or departure time choice model. This is particularly important when considering a dynamic problem (where tolls vary over time) such as to properly account for how tolls impact the temporal distribution of demand.

The simulator accounts for time-varying network demand, link demand, and link states. The analytical model used to build the metamodel is a stationary model, which does not account for temporal variations in demand and link states. The extension of these ideas to address a dynamic problem (where tolls vary across time intervals) will require the use of an analytical model that captures these temporal variations. A straightforward approach would be to formulate a time-dependent analytical traffic model. The main challenge would then be to derive a sufficiently tractable analytical formulation. A different approach, which has recently been proposed for a class of high-dimensional SO problems (OD calibration problems), is to use one stationary model per time interval and to account for the temporal variations through the metamodel parameters (Osorio 2019). The extension of these ideas for dynamic toll optimization is part of ongoing work.

The approach of this paper enables a detailed, computationally inefficient, simulation-based traffic model to be used to tackle, in a computationally efficient way, a toll optimization. Computational efficiency is achieved by combining, at every iteration of the algorithm, information from the simulation-based model with information from a simplified analytical and computationally efficient traffic model. In this paper, the analytical model is specified as a system of differentiable nonlinear equations. The use of other analytical traffic models, and the study of their corresponding accuracy-efficiency tradeoffs, is of interest for future work. This could lead to new and interesting challenges in the formulation and solution of the corresponding metamodel optimization problems.

An interesting extension is to account for a multiclass setting with different classes (i.e., types) of users.

This could allow for a more realistic representation of demand with different types of travelers (e.g., commuters, infrequent travelers) with different values of time or willingness to pay. In particular, there is a recent interest in the field for the design of personalized tolls (i.e., there is a toll for each user class), following the ideas of personalized mobility services (Danaf et al. 2019).

In this paper, the route choice sets of both the simulator and the analytical traffic models are predetermined and computed offline. In the case where the traffic simulator considers an endogenous route choice generation (e.g., the route choice set is computed on the fly prior to each network loading operation), it is of interest to formulate a metamodel framework where the route choice set of the analytical model is also endogenous. The main methodological challenge here is to analytically account for the impact of simultaneous changes in both the decision vector and the route choice set on the analytical objective function and to do so in a scalable way such that it can used for networks with high-dimensional route choice sets.

The approach of this paper enables a detailed, computationally inefficient, simulation-based traffic model to be used to tackle, in a computationally efficient way, a toll optimization. Computational efficiency is achieved by combining, at every iteration of the algorithm, information from the simulation-based model with information from a simplified analytical and computationally efficient traffic model. In this paper, the analytical model is specified as a system of differentiable nonlinear equations. The use of other analytical traffic models, and the study of their corresponding accuracy-efficiency tradeoffs, is of interest for future work. This could lead to new and interesting challenges in the formulation and solution of the corresponding metamodel optimization problems.

Acknowledgments

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors acknowledge the support of the Intelligent Transportation Systems (ITS) Laboratory both at MIT and SMART (Singapore-MIT Alliance for Research and Technology). Special thanks to Dr. Ravi Seshadri for his help in using the Singapore network model, which was developed and calibrated by his DynaMIT team at SMART.

References

Albalate D, Bel G, Fageda X (2009) Privatization and regulatory reform of toll motorways in europe. *Internat. J. Policy Admin. Institutions* 22(2):295–318.

Chen X, Osorio C, Santos BF (2019) Simulation-based travel time reliable signal control. *Transportation Sci.* 53(2):523–544.

Chen X, Zhang L, He X, Xiong C, Zhu Z (2018) Simulation-based pricing optimization for improving network-wide travel time reliability. *Transportmetrica A: Transportation Sci.* 14(1-2):155–176.

Chen XM, Xiong C, He X, Zhu Z, Zhang L (2016) Time-of-day vehicle mileage fees for congestion mitigation and revenue generation: A simulation-based optimization method and its real-world application. *Transportation Res. Part C: Emerging Tech.* 63:71–95.

Chong L, Osorio C (2018) A simulation-based optimization algorithm for dynamic large-scale urban transportation problems. *Transportation Sci.* 52(3):637–656.

Corthout R, Himpe W, Viti F, Frederix R, Tampere CM (2014) Improving the efficiency of repeated dynamic network loading through marginal simulation. *Transportation Res. Part C: Emerging Tech.* 41:90–109.

Danaf M, Becker F, Song X, Atasoy B, Ben-Akiva M (2019) Online discrete choice models: Applications in personalized recommendations. *Decision Support Systems* 119:35–45.

de Palma A, Lindsey L (2011) Traffic congestion pricing methodologies and technologies. *Transportation Res. Part C: Emerging Tech.* 19(6):1377–1399.

Dong J, Mahmassani HS, Erdogan S, Lu CC (2011) State-dependent pricing for real-time freeway management: Anticipatory vs. reactive strategies. *Transportation Res. Part C: Emerging Tech.* 19(4):644–657.

Dong NA, Eckman DJ, Poloczek M, Zhao X, Henderson SG (2017) Comparing the finite-time performance of simulation-optimization algorithms. Chan WKV, D'Ambrogio A, Zacharewicz G, Mustafee N, Wainer G, Page E, eds. *Proc. 2017 Winter Simulation Conf* (IEEE, Piscataway, NJ).

Google Maps (2017) *Singapore Expressway Network*.

Gupta S, Seshadri R, Atasoy B, Prakash AA, Fereira F, Tan G, Ben-Akiva M (2020) Real-time predictive control strategy optimization. *Transportation Res. Record* 2674(3):1–11.

Halvorson R, Nookala M, Buckeye KR (2011) High occupancy toll lane innovations: I-394 mnpass. *Transportation Res. Board Annual Meeting, Washington DC*.

Han L, Wang DZW, Lo HK, Zhu C, Cai X (2017) Discrete-time day-to-day dynamic congestion pricing scheme considering multiple equilibria. *Transportation Res. Part B: Methodology* 104:1–16.

Hassan A, Abdelghany K, Semple J (2013) Dynamic road pricing for revenue maximization: Modeling framework and solution methodology. *Transportation Res. Rec.* (2345):100–108.

Jang K, Chung K, Yeo H (2014) A dynamic pricing strategy for high occupancy toll lanes. *Transportation Res. Part A: Policy Practice* 67(C):69–80.

Land Transport Authority (2019) *Erp Rates*. Accessed August 26, 2019, https://www.onemotoring.com.sg/content/onemotoring/en/on_the_roads/ERP_Rates.html.

Liu Z, Wang S, Zhou B, Cheng Q (2017) Robust optimization of distance-based tolls in a network considering stochastic day to day dynamics. *Transportation Res. Part C: Emerging Tech.* 79:58–72.

Lou Y, Yin Y, Laval JA (2011) Optimal dynamic pricing strategies for high-occupancy/toll lanes. *Transportation Res. Part C: Emerging Tech.* 19(1):64–74.

Mathworks, Inc. (2016) *Global Optimization Toolbox User's Guide Matlab (R2016b)* (The MathWorks, Natick, MA).

Osorio C (2019) Dynamic origin-destination matrix calibration for large-scale network simulators. *Transportation Res. Part C: Emerging Tech.* 98:186–206.

Osorio C, Bierlaire M (2013) A simulation-based optimization framework for urban transportation problems. *Oper. Res.* 61(6):1333–1345.

Osorio C, Chong L (2015) A computationally efficient simulation-based optimization algorithm for large-scale urban transportation. *Transportation Sci.* 49(3):623–636.

Osorio C, Nanduri K (2015) Energy-efficient urban traffic management: a microscopic simulation-based approach. *Transportation Sci.* 49(3):637–651.

Osorio C, Selvam K (2017) Simulation-based optimization: achieving computational efficiency through the use of multiple simulators. *Transportation Sci.* 51(2):395–411.

Saleh W, Sammer G (2009) Travel demand management and road user pricing: Success, failure and feasibility. Saleh W, Sammer G, eds. *Travel Demand Management and Road User Pricing: Success, Failure and Feasibility* (Routledge, New York), 1–9.

Seik FT (2000) An advanced demand management instrument in urban transport: Electronic road pricing in Singapore. *Cities* 17:33–45.

Tay T, Osorio C (2017) An efficient sampling method for stochastic simulation-based transportation optimization. *First INFORMS Transportation Sci. Logistics (TSL) Conf., Chicago.*

Toledo T, Mansour O, Haddad J (2015) Simulation-based optimization of hot lane tolls. *Transportation Res. Procedia* 6:189–197.

Vu V, Hashemi H, Seshadri R, Gupta S, Tan G, Prakash AA, Ben-Akiva M (2018) Predictive distance-based toll optimization under elastic demand for real-time traffic management. *Transportation Res. Board Annual Meeting, Washington DC.*

Wang W, Wan H, Chang KH (2016) Randomized block coordinate descendant STRONG for large-scale stochastic optimization. *2016 Winter Simulation Conf.* (IEEE), 614–625.

Xu J, Nelson BL, Hong JL (2013) An adaptive hyperbox algorithm for high-dimensional discrete optimization via simulation problems. *INFORMS J. Comput.* 25(1):133–146.

Zhang Y, Atasoy B, Ben-Akiva M (2018) Calibration and optimization for adaptive toll pricing. *Transportation Res. Board Annual Meeting, Washington DC.*

Zhang C, Osorio C, Flötteröd G (2017) Efficient calibration techniques for large-scale traffic simulators. *Transportation Res. Part B: Methodology* 97:214–239.

Zheng N, Rérat G, Geroliminis N (2016) Time-dependent area-based pricing for multimodal systems with heterogeneous users in an agent-based environment. *Transportation Res. Part B: Methodology* 62:133–148.