Transportation Research Part C 62 (2016) 133–148

Contents lists available at ScienceDirect

Transportation Research Part C

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

Time-dependent area-based pricing for multimodal systems with heterogeneous users in an agent-based environment

Nan Zheng, Guillaume Rérat, Nikolas Geroliminis *

Urban Transport Systems Laboratory, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

ARTICLE INFO

Article history:
Received 8 May 2015
Received in revised form 19 October 2015
Accepted 25 October 2015
Available online 28 November 2015

Keywords:
Congestion pricing
Multimodal
User heterogeneity
Macroscopic Fundamental Diagram
User adaptation
Agent-based

ABSTRACT

In this paper, we investigate an area-based pricing scheme for congested multimodal urban networks with the consideration of user heterogeneity. We propose a time-dependent pricing scheme where the tolls are iteratively adjusted through a Proportional–Integral type feedback controller, based on the level of vehicular traffic congestion and traveler's behavioral adaptation to the cost of pricing. The level of congestion is described at the network level by a Macroscopic Fundamental Diagram, which has been recently applied to develop network-level traffic management strategies. Within this dynamic congestion pricing scheme, we differentiate two groups of users with respect to their value-of-time (which related to income levels). We then integrate incentives, such as improving public transport services or return part of the toll to some users, to motivate mode shift and increase the efficiency of pricing and to attain equitable savings for all users. A case study of a medium size network is carried out using an agent-based simulator. The developed pricing scheme demonstrates high efficiency in congestion reduction. Comparing to pricing schemes that utilize similar control mechanisms in literature which do not treat the adaptivity of users, the proposed pricing scheme shows higher flexibility in toll adjustment and a smooth behavioral stabilization in long-term operation. Significant differences in behavioral responses are found between the two user groups, highlighting the importance of equity treatment in the design of congestion pricing schemes. By integrating incentive programs for public transport using the collected toll revenue, more efficient pricing strategies can be developed where savings in travel time outweigh the cost of pricing, achieving substantial welfare gain.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Research in transport economics have proposed road pricing as an effective policy to relieve traffic congestion in cities for many years. By charging road users the external costs they create, congestion pricing aims to trigger travel behavior changes (e.g. mode shift from cars to buses or departure time shift to outside of peak-hour) such that congestion is avoided. Despite the vast literature, a small number of cities have actually implemented congestion tolls due to social disagreement, political issues and myopic treatment of other modes of transport. Traffic mobility in multimodal systems is inherently a distributed and interconnected process, which should be modeled and managed as a whole to improve the global operational efficiency.

* Corresponding author.
E-mail addresses: nan.zheng@epfl.ch (N. Zheng), guillaume.rerat@epfl.ch (G. Rérat), nikolas.geroliminis@epfl.ch (N. Geroliminis).

http://dx.doi.org/10.1016/j.trc.2015.10.015
0968-090X/© 2015 Elsevier Ltd. All rights reserved.

The classical approach for determining pricing is based on economic theories. In the literature, there exists an extensive body of works on the first- and second-best pricing models. A comprehensive summary of these models can be found in Yang and Huang (2005) and de Palma and Lindsey (2011). Over the years, researchers have made significant effort in extending this type of models (either of a bottleneck or a small network) for studying the impact of pricing. Representative examples of this direction can be found in Verhoef (2002), Arnott (2007), and elsewhere, where Vikrey's bottleneck model (Vickrey, 1969) form the basis of their analysis; or in Lu et al. (2008), Lou et al. (2010) and Wu et al. (2011), where user equilibria of route choices were derived, given that the cost of pricing is taken into account. The treatment of user heterogeneity and equity should not be ignored when discussing congestion pricing. Regarding this research direction, similar types of models were proposed to capture the behavioral difference among travelers with have different value-of-times (see for instance in Lu et al., 2008; van den Berg and Verhoef, 2011; Qian and Zhang, 2013; Tian et al., 2013). Other well-established works on the design of equitable pricing schemes for multimodal networks include for example, Yang and Zhang (2002) and Yin and Yang (2004) who developed optimal pricing with social and spatial equity constraints. Others employed Pareto-improving approaches, such that all users are not worse off in the presence of pricing, see for example Lawphongpanich and Yin (2010), Nie and Liu (2010) and Xiao et al. (2013). Wu et al. (2012) and Xiao et al. (2013) share similar research interest with this paper, where the distributional effect of pricing on different income groups is captured and income-based pricing schemes are discussed. Zhu et al. (2013) developed a theoretical framework to address equity issues, though they did not touch directly how an equitable pricing scheme should be theoretically designed.

Existing traffic models in transport economics pose severe theoretical and empirical limitations in realistic applications. The main reason is because they employ link travel cost functions, which do not accurately describe the intra-day traffic dynamics and relate them to urban-scale network characteristics in a way that is computationally tractable and consistent with the physical properties of traffic. This failure constrains the ability of economic models to support efficient developments for network traffic management and diminish congestion externalities (Tsekeris and Geroliminis, 2013). With respect to the treatment of congestion dynamics (utilized traffic models), most of the aforementioned studies assume a steady-state traffic condition or demand-type of capacity-supply functions. Such models ignore the fact that the level of congestion is not a memory-less function of demand at a given time, but dependent on the history of the system, i.e. the same demand profile can influence differently the system if this was uncongested or congested at the current state (Geroliminis and Levinson, 2009).

With respect to the application level of the pricing schemes, pricing based on individual links is extremely difficult to implement in practice, and it is computationally complex for large-scale networks (Verhoef, 2002). With respect to the consideration of user heterogeneity, few studies quantitatively analyze the impact of pricing on different types of users, e.g. differences in mode shift behavior or distributional effect (as pointed out by Eliasson and Mattsson, 2006), while designing a pricing scheme capable of balancing the gain of all users is even rarely mentioned in the literature. While with respect to multimodality, alternative modes are not taken into account in most approaches. Integrating multimodality consideration in pricing is challenging and reserves further attention.

While the classical approaches focus on the link- or corridor-level of application, existing pricing schemes for macroscopic (large-scale) level have similar ambiguity. For example, in Maruyama and Sumalee (2007) and Zhang et al. (2008) traffic conditions are considered stationary; whereas in the well-known field implementation of congestion pricing, the Singapore scheme, it introduced a dynamic area-based type pricing system where the toll rates are time-dependent and adjusted based on regular surveys on travel speed (Liu et al., 2013). The valuable experience of the Singapore case significantly advanced the knowledge on congestion pricing, and had inspired the developments of speed-based toll schemes where a pre-defined optimal speed range serves as the base for toll optimization. According to the same reference mentioned above, the speed-based toll is regulated such that traffic flow rates can maintain at a speed ranging from 20 km/h to 30 km/h. Though it sounds reasonable, the issue is that traffic states can vary significantly at this speed range (see for example in the empirically observed fundamental diagram, e.g. in Geroliminis and Daganzo, 2008), therefore the determined pricing rates may not guarantee an optimum. Regarding the pricing schemes of other successful field implementations, for instance in Stockholm and London, the underlying method for determining the prices is unfortunately unavailable and remains as a black-box to the researchers.

To develop an effective congestion pricing scheme for urban networks, the aforementioned inadequacies need to be treated. Integrating a proper traffic model that captures congestion dynamics is critical in designing the optimal prices. It is even more challenging as the current urban management requires understanding at the network level. Recent findings on the Macroscopic Fundamental Diagram (MFD) shed light into this direction. The idea of macroscopic traffic model for car-only urban networks was initially proposed in Godfrey (1969) and followed in Mahmassani et al. (1987) and Daganzo (2007). The demonstration of the existence of the MFD with dynamic features from field data was firstly reported in Geroliminis and Daganzo (2008), showing that urban single-mode regions exhibit an MFD relating space-mean network flow (traffic throughput) to network density (traffic state). It has been showed by the same reference that (i) even though the flow-density plots for individual links in the network exhibit considerable scatter, the scatter nearly disappears once data of the individual links are spatially aggregated for the entire network, and (ii) the MFD is a property of the network itself (infrastructure and control) and not very sensitive to different demand patterns. An interested reader could refer to Yildirimoglu and Geroliminis (2014) and Leclercq et al. (2014) for a review of recent developments in MFD. Given the MFD of a network, monitoring and predicting the evolution of congestion is possible through the existing data technologies with fixed-location and mobile sensors (loop detectors and GPS of probe vehicles). It should be noted that the spread of

congestion may not be homogeneous within large-scale networks. If the MFD of a network exhibits high spatial or modal usage heterogeneity, a solution is to partition such network into regions with homogenous congestion distribution that each partitioned-region has a low-scatter MFD (Ji and Geroliminis, 2012) and a higher degree of connectivity and compactness (on-going work investigates further this direction Saeedmanesh and Geroliminis, 2015). A modeling framework has been developed to capture how vehicles passing such multi-region networks with MFD-representation (Yildirimoglu et al., 2015). Furthermore, latest works extend the single-mode MFD to a bi-modal where cars and buses share the same infrastructure and look at passenger flow dynamics in addition to vehicular dynamics (Zheng and Geroliminis, 2013; Geroliminis et al., 2014; Chiabaut et al., 2014; Chiabaut, 2015).

One of the most important contributions of the MFD and bi-modal MFD modes is that they enable the development of traffic management strategies to optimize bi-modal urban networks. Research effort has been made by many researchers, e.g. perimeter flow control for single-mode networks (Aboudolas and Geroliminis, 2013; Keyvan-Ekbatani et al., 2015; Haddad, 2015), and multimodal networks (Ampountolas et al., 2014). In addition to online traffic control, studies have shown promising results on demand management with the MFD, as well. Gonzales and Daganzo (2012) examined system optimum solutions using MFD for a transport system with cars and public transit share spaces for the morning commute problem, whereas Gonzales and Daganzo (2013) provided solutions for combined morning and evening peaks problems. The first work of MFD-based pricing investigated a flat-rate cordon pricing during the peak hour (Zheng et al., 2012). This work showed that an MFD-based pricing scheme succeeds not only in congestion reduction, but also high efficiency where the total travel time savings outweigh the total toll paid. In this work, travelers could only change their time of departure and route but mode choice is not an option. A similar approach was later applied in Simoni et al. (2015) who derived optimal pricing using MFD and marginal cost theory, without investigating the impact of pricing on multimodality either. Combining dynamic pricing with dynamic allocation of road space between multiple modes (e.g. changing the fraction of roads dedicated to buses) using MFD modeling has been shown to be an efficient traffic management scheme (Zheng and Geroliminis, 2013). Nevertheless, the aforementioned work performs the optimization by assuming perfect knowledge of how users adapt due to congestion and pricing through an MFD model. Investigating dynamic traffic management schemes where the model has a different level of abstraction from the considered reality (plant is the appropriate terminology in control literature) can shed more light for the applicability of these schemes in the real world. In this work, while the pricing is developed based on an aggregated MFD model, the implementation is done in a more detailed model in an agent-based environment, where information for the actions and the individual behavior of users is unknown during the optimization phase. If such an approach shows significant improvements in terms of congestion, it can be a promising and convincing methodology to be applied in real life. While many studies have showed that a time-dependent toll (e.g. based on congestion level) can reduce congestion more efficiently, a network time-dependent area-based pricing building on the MFD representation of congestion for multi-modal networks deserves investigation.

To develop efficient pricing schemes that allow higher user acceptance, we consider the following objectives in our current work. (i) integrate the behavioral adaptation of the users during pricing (ii) promote a multimodal solution combined with pricing, to maximize the efficiency of pricing, and (iii) recognize the heterogeneous effect of pricing on users, to design pricing strategies with equity. We will show in this paper that proper treatments, such as incorporating a value-of-time based pricing rates or redistributing toll revenue to improve public transport service, can significantly enlarge the efficiency of the pricing and the sustainability of urban mobility. In particular, we reveal that considering the effect of pricing on multimodal mobility improves remarkably the system performance, which has not been studied thoroughly. Questions that arise are how pricing rates should be adjusted when congestion conditions change due to user's behavioral changes? What are the impacts of different types of incentive programs for using public transport (PT) on the performance of congestion pricing? It has to be highlighted that despite the vast literature in pricing, field tests are quite limited and this is among other reasons of user acceptability. Thus, providing incentives to other modes of transport (e.g. return a fraction of the tolls paid to users that switched to public transport mode) can make such policies more attractive for real cases.

In this paper, we propose a time-dependent area-based congestion pricing scheme for multimodal urban network with heterogeneous user groups. Two groups of users are differentiated in our approach by their individual VOTs. We determine time-dependent tolls based on the level of congestion of the network represented by the MFD, and user's behavioral adaptation to the tolls captured by the evolution of congestion between the adjustments of pricing rates. An agent-based model, which preferably ensures a realistic reproduction of travel behavior of users, is utilized for carrying case studies. Only aggregated information is collected from the model, and passed to the higher level traffic model (the MFD) which is utilized during the optimization phase to determine the optimal toll. We investigate the performance of pricing strategies that integrates public transport incentives into the pricing schemes, and we analyze the distributional effect of pricing to the two different user groups. Note that the pricing strategy is based on a feedback control approach, which utilizes observable information when determining the prices. There is no behavioral model integrated during pricing, but the change in behavior from day to day (or period to period) influences the value of price.

The rest of the paper is organized as follows. Section 2 describes in detail the methodology for treating user heterogeneity and the control algorithm of pricing. Section 3 presents the case study results of the Sioux Fall network, and compares the performances of four pricing strategies which differ in the treatment on user adaptation and the integration of multimodality in pricing. Dedicated analysis and discussion are given in Section 4, regarding the behavioral differences and distributional effect between users having heterogeneous values-of-time. Concluding remarks are provided at the end.

2. Methodology

This section is divided into three parts. Section 2.1 describes the properties of the employed agent-based model, the simulation process, and the proposed treatment on user heterogeneity to integrate into the simulation model. Section 2.2 concerns the core concept of an MFD- and user-adaptation based pricing scheme, providing the algorithmic steps to reach the desired mobility level and detailed explanations on pricing adjustment. Having understood the model and the pricing control mechanism, Section 2.3 introduces two pricing strategies that combine a time-dependent area-based pricing and incentives of promoting the usage of public transport.

2.1. Agent-based simulation and user heterogeneity

In this work, while the pricing is developed based on an aggregated MFD model, the implementation is done in a more detailed agent-based environment, where information for the individual behavior of users is unknown during the optimization. We consider this approach to have a high degree of realism, given that detailed behavioral data is difficult to be obtained in real time and even more difficult to be applied in an optimization process. It is also known that detailed simulations contain a large number of assumptions that do not necessarily represent reality with a high degree of accuracy, especially with respect to the dynamic development and evolution of congestion. Thus, our intension is to test a feedback control framework to improve mobility of multi-modal systems based on aggregated network level modeling, in an environment with significantly higher degrees of freedom and complexity than the one the model is not aware of.

The agent-based simulation model MATSim is utilized (Meister et al., 2010). The MATSim model has been widely applied for travel behavior studies in transportation systems and land use analysis. It integrates activity-based demand generation with dynamic traffic assignment. Demand generation is embedded in a concept of daily activity sequence from which the need for transport is derived. In the context of activity-based demand, the entire activity plan (mode choice, departure time choice, route choice and the activity sequence) is the unit of decision to iterate travel dynamics. Random utility theory is applied to generate the plans of daily activities. Each agent (traveler) in the simulation is assigned with different utility values when performing different activities. A typical utility function consists of four items for an agent performing her daily plan: denote $U_{i,act}$ the score performing activity i, $U_{i,travel}$ the score of traveling to activity i, $U_{i,wait}$ a penalty for waiting instead of performing activity i, and $U_{i,short}$ a penalty for performing activity i for a too short duration, where $i \in \{1, \ldots, n\}$ is the number of planned activities. The existing model does not distinguish the difference of these utility terms when perform exactly the same activity. For example, a travel time of 10 min would be simply transformed to a utility value of $U_{i,travel}$ 3CHF, if assuming a value-of-time is 18 CHF/h.

Now let us consider the existence of heterogeneity in the agent population. Agents are grouped with respect to value-of-time (VOT). Denote j as the group index Eq. (1) displays the utility calculation for user group j performing daily plans:

$$U^j_{plan} = \sum_{i=1:n} U^j_{i,act} + U^j_{i,travel} + U^j_{i,wait} + U^j_{i,short} \tag{1}$$

Note that in the current work, the influence of VOT is mainly reflected by the physical travel cost term $U^j_{i,travel}$. Nevertheless it is easy to incorporate this treatment to other terms, such as embedding earliness and lateness penalty into $U^j_{i,act}$. These properties provide more realistic modeling of the reaction of the agents towards any traffic management strategy, and thus can help traffic engineers develop more reliable congestion pricing schemes. Furthermore, we assume in this work that VOT varies across individual income, which is line with the empirical findings by Börjesson and Eliasson (2014). Defining user heterogeneity in this manner allows us to later design equitable pricing strategies among for example the "rich" and the "poor" people. Other studies indicate that difference in VOT also exists in mode choice preference (such as Börjesson et al., 2014).

As for the simulation, the process can be summarized as follows. Each agent has an activity plan to perform her daily events. The selected activity plans are simulated along the timeline in the model representation of the physical world. A First-In-First-Out queue model is embedded as the traffic model, with pre-defined road capacity and some treatment of shockwave between vehicles traveling backwards at constant speed in the case of discharging traffic jam. The executed activity plans are evaluated with a measure of utility as mentioned above. Based on the evaluation result, certain amount of the agents (by default 10%) carries out a re-planning strategy such as changing mode choice or departure time. Smaller values will significantly increase convergence times, while larger values might create strong oscillatory behavior. New plans then are created and added into the memory of the agents. Agents afterwards decide to execute either the new plan or choose one of the existing plans, preferably the ones with the highest evaluation score. This procedure is iterated via a day-to-day learning process, until agent-based stochastic user equilibrium is achieved (Nagel and Flötteröd, 2009). Under such equilibrium, not only travel behavior is stabilized, but also the utility of the entire daily activities, U^j_{plan}, is indirectly optimized. Please note that as it is complicated to achieve the classical stochastic user equilibrium for systems with such multi-dimension behaviors. Trade-off needs to be made between computational cost of the iterations (e.g. a full-scale one-day simulation of the Swiss city Zurich takes eight hours on a high-performance-computer) and the magnitude of behavioral convergence. For detailed information on this subject or MATSim, the readers may refer to Nagel and Flötteröd (2009) and Meister et al. (2010).

2.2. MFD-based pricing scheme with user adaptation

The physical modeling tool, which is integrated in time-dependent pricing is the Macroscopic Fundamental Diagram (MFD) of urban traffic, which provides for some network regions with a small heterogeneity in the spatial distribution of congestion, a unimodal, low-scatter relationship between network vehicle density (veh/km) and network space-mean flow (veh/h). The MFD offers an analytically tractable and parsimonious approach for modeling congestion with dynamic features and is relevant to pricing-oriented policies to reduce congestion. The MFD approach permits the analysis of a set of second-best strategies, which could not be investigated in classical economic models that fail to include hypercongestion or are intractable for large-scale networks. Such measures include the use of advanced monitoring technology for metering of access or perimeter control (Aboudolas and Geroliminis, 2013; Ramezani et al., 2015; Keyvan-Ekbatani et al., 2015 and others), the reallocation of the existing network capacity and land use among urban zones (Tsekeris and Geroliminis, 2013) and area-based pricing (Zheng et al., 2012). The main logic of the strategies is that they try to decrease the inflow in regions with points in the decreasing part of an MFD.

The MFD can be obtained via analytical approximations with variational theory (such as proposed in Geroliminis and Boyacı (2012), Leclercq and Geroliminis (2013), or observed with real data from multiple sensors Leclercq et al. (2014), Ortigosa et al. (2014), Ji et al. (2015) and others). If network density is regularly measured, the critical density K_{cr} beyond which oversaturation accumulates in the network and congestion grows, can be identified. Knowing K_{cr} is important and extremely helpful for traffic management, as the general principle of an MFD-based strategy aims at operating a network at traffic states where density is no higher than K_{cr}. Once the network accumulation exceeds the critical density K_{cr}, control actions will then be activated to guide the network to recover to states around K_{cr}. This type of actions in control theory literature have been addressed in transportation science mainly with integration of feedback control, and has been widely applied in traffic regulation on real freeways (Papageorgiou and Kotsialos, 2010) and flow restriction at the perimeter of urban networks (see previous paragraph for references), or mixed arterial-freeway networks for simple networks (Haddad et al., 2013). Developing pricing strategies with similar control tools have an additional challenge. While in traffic control, the effect of feedback is immediate and drivers might not have the ability to adapt (they might change only their route), pricing could make users to make behavioral changes, e.g. change their departure time from the origin or change mode of transport. Considering this adaptation in the development and design of efficient pricing schemes could be challenging, as the control action might not have the expected results. While large amount of research efforts have been put to develop traffic control strategies based on the concept of the MFD, adaptation of travelers to these changes have not been considered.

We attempt to build our pricing scheme to cover both aspects mentioned above. Given that the multi-dimension behavioral adaptation is quite difficult to model in an analytical way, we consider that an agent-based framework influences their decisions in an unknown way to the system operators (traffic manager). Nevertheless, the operators can monitor the aggregated network performance through vehicle accumulations (or densities). The monitoring of vehicle accumulation can also help identify spatial heterogeneity in congestion within a network. If the distribution of congestion exhibits strong heterogeneities, a partitioning algorithm (for example Ji and Geroliminis, 2012) should be applied to determine the area for pricing implementation. A coordinated multi-area pricing scheme will be reported as a future research direction. In this work a time-dependent area-based pricing scheme is applied. This means that any trip traveling in the area under consideration, should pay the specified toll independently of the distance traveled. Trip-based tolls have been reported in the literature (Levinson, 2010) as possibly more equitable, but also more difficult to implement.

A successful pricing scheme should consider that travelers adapt their behavior through a day-to-day learning process and the aggregated effect of this behavioral adaptation might make the current pricing rate inefficient. The proposed pricing control scheme takes this effect into account and adjusts the prices after some time periods. Recall that the agent-based simulation performs iterations such that a system convergence point is reached. If prices change in every iteration, then the travelers do not have enough time to evaluate their decisions and convergence might take too long because the learning process of the drivers based on others' choices is slower due to the intervention in the system. To be able to reflect the impact on traffic performance when a certain pricing rate is applied, this pricing rate will remain unchanged during a required number of iterations where the system is close to convergence. In our results, the system reaches close-to-equilibrium with non-changing decisions for the travelers after around 50 iterations. We therefore take the system performance of the 50th iteration as the representation of the permanent effect of the current pricing, and adjust the pricing such that less congestion would be expected after the users re-adapt in the new pricing. A similar concept for pricing update is implemented in San Francisco, where on-street parking prices are re-evaluated once per month (Pierce and Shoup, 2013).

Summarizing the discussion above, an illustration of this pricing scheme is displayed in Fig. 1. Detailed description of the algorithmic steps is provided as the follows:

(i) Given an initial scenario without pricing, identify the congested area of the city where pricing will be applied and obtain the shape of the MFD. In a real environment, this can be obtained from multi-sensor data and a partitioning algorithm. For this paper, we utilize an agent-based simulation. The critical value of density that maximizes the network flow, K_{cr} is estimated.

(ii) Observe the time series of network and identify the time intervals when the network density is higher than K_{cr} and congestion occurs. Determine initial tolls by applying only the term $c_1(K_k(t) - K_{cr})$ of Eq. (2).

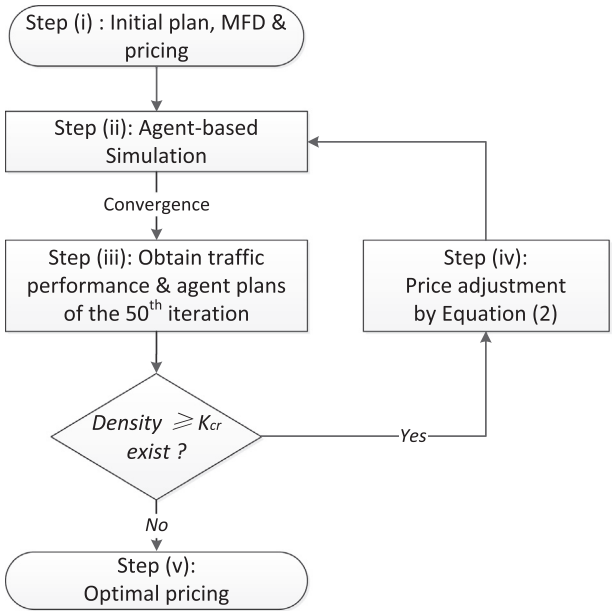

Fig. 1. An illustration of the optimization algorithm of pricing.

(iii) Let the system operate for a number of days with the determined tolls, until users learn and adapt and convergence is reached (density time-series does not change significantly from day to day). In this paper, we run the agent-based simulation with the updated toll for 50 iterations when system convergence is closely reached.[1]

(iv) Obtain the resultant traffic performance (final density time-series). Check if the network experiences states that exceed K_{cr}. If there are still time intervals with congested states, apply a Proportional-integral type controller to adjust the tolls, which is defined by Eq. (2). Repeat from step (ii).

(v) If $K_k(t) < K_{cr}$ for all time intervals, finish price update and obtain the final toll.

Let us now focus on the equation for adjusting the pricing. Denote k, the k-th adjustment of the pricing. The toll at the (k+ 1)-th toll adjustment for time interval t, $Toll_{k+1}(t)$, is proportional to (i) the magnitude that the average network density $K_k(t)$ exceeds K_{cr} and (ii) the difference between the resultant densities under the current pricing $K_k(t)$ and the previous one $K_{k-1}(t)$. Recall that K_{cr} is the point in the MFD, where the maximum network production is reached and network production decreases if $K_k(t) > K_{cr}$). A slightly smaller value is chosen, e.g. 0.95 K_{cr}, to facilitate stability of the system around the set point.

Eq. (2) below is the pricing controller, where c_1 and c_2 are control gain parameters:

$$Toll_{k+1}(t) = Toll_k(t) + c_1(K_k(t) - K_{cr}) + c_2(K_k(t) - K_{k-1}(t)) \tag{2}$$

This equation employs a Proportional–Integral (PI) type controller which is a classic feedback control mechanism. It was demonstrated that dynamic systems with well-defined properties (e.g. small errors in the traffic states) can be stabilized with a feedback strategy to a desired state (Papageorgiou and Kotsialos, 2010). Similar dynamic control of pricing was applied to the managed lanes (single-link level) with consideration on user's willingness to pay (Yin and Lou, 2009). Comparing to the pricing schemes developed in Zheng et al. (2012) and Zheng and Geroliminis (2013), the proposed scheme gives higher flexibility in toll adjustment based on user's adaptation to the toll. In particular, this scheme allows reduction of toll if $K_k(t)$ is smaller than $K_{k-1}(t)$ which benefits the users from not being overcharged. Parameters c_1 and c_2 have constant and positive values. The values are chosen offline via a trial-and-error process to avoid oscillations (analytical solution would be impossible to obtain for such a system with complex dynamics). We would emphasize that this pricing strategy is robust to moderate parameter changes and has fast and global convergence. A sensitivity analysis on c_1 and c_2 will be provided at the final part of the paper. Alternatively, price can be determined by solving optimization problems in real-time. An example can be found in Lou et al. (2011) where prices are optimized to maximize the maximum flow of a freeway section meanwhile maintain a free-flow condition for the high-occupancy-toll lane.

Note that a PI controller is applied for every time t independently of other times. The feedback is provided by iteration k, which describes the adaptation of drivers from iteration to iteration (or day to day in the real world). Despite the fact that

[1] In other networks or real environments the time to convergence might vary. By observing how the system evolves with time, close-to-convergence conditions can be observed with real data.

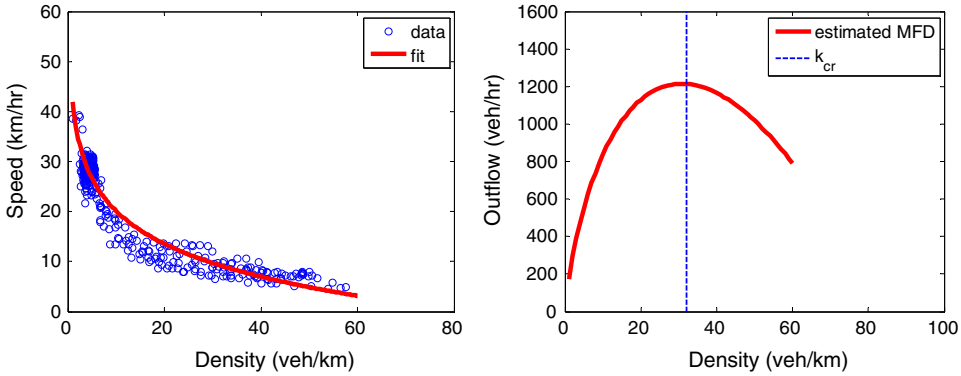

Fig. 2. The MFD of the study site: (a-down left) network speed-density data scatter and (b-down right) the fitted network flow-density plot with indication of the critical density K_{cr}.

these controllers operate independently, it is found that the resultant pricing scheme is smooth with respect to time t. The reason is that the space-mean density of the network $K_k(t)$ evolves smoothly over time t while correlation exists between the level of congestion between the current and its previous time step t. Thus, the pricing scheme is based on the iteration-to-iteration (or day-to-day) dynamic behavior and not in within-day dynamics, i.e. it is not a real-time pricing scheme. If such a pricing scheme was developed at a link basis, it would be expected that results would have quite oscillatory behavior due to the non-smooth dynamics of traffic for links (compared to spatial aggregations at network levels as described in the MFD literature). Note that the density of the network at the final toll, should not be significantly smaller than K_{cr} during the peak period. For example if the set-point of the controller is chosen at 0.5 K_{cr}, the capacity of the network is underutilized and the system will operate at higher speed, but much lower flows. Note also that, the choice of K_{cr} and the final prices can be seen as policy factors influencing the service level of general mobility. One may argue that if lower values are chosen, the toll is very strict and over-charging might occur as the system operates in a state less than its capacity. Nevertheless, this does not necessarily indicate over-charging unless the total toll paid outweighs the total savings in travel time. System operator can choose the desired values in a way to maximize traffic flow and as a result to transfer the highest possible number of passengers with small delays. The value of toll which results in points in the MFD just without the decreasing branch (a graphical illustration can be found in Fig. 2) can be set as the lower bound, while the one which results the total toll paid just equal to total travel time savings can be set as the upper bound in order to avoid over-charging. This higher toll will push the system to operate at a smaller than the maximum flow during the toll period, but with higher average speed and some potential savings in travel delay. This also implies that an operator can utilize a slightly smaller value to guarantee a more reliable operation, e.g. capable of buffering if there is sudden disturbance from demand, as flow decreases (see for example Geroliminis and Levinson, 2009).

2.3. Pricing strategies with accessibility improvement and incentive on public transport

Road traffic networks and management systems constitute a system of systems (SoS) with complex behaviors of various modes of transport competing for the same space. A traditional approach of decentralized control design for traffic management is not appropriate for heterogeneously loaded networks with multiple modes of transport and high congestion levels (Haddad et al., 2013). Lack of collaboration between operators of different systems can result in sub-optimal and less efficient conditions because it does not consider interactions and conflicts between different modes. Traffic systems should be managed as a whole to improve their global operational efficiency while avoiding conflicting decisions. The proposed MFD-based traffic management strategies thus should serve for this sustainable and global mobility goal. This requires the consideration of alternative modes in the design process of pricing, understanding the impact of pricing on the operators and the users of traffic systems, and development of integrated and smart pricing schemes that motivate user acceptability and behavioral changes, as well as reaching equitable savings for all users.

We propose in this paper to simultaneously treat multimodality and equity with congestion pricing. Recently, several works (Zheng et al., 2012; Simoni et al., 2015) are devoted to develop effective congestion pricing schemes similar to the one of Section 2.2, however without treating multimodality or user heterogeneity. Operating at a decent level of public transport (PT) services is important. If public transport is not attractive and mode shift is limited, congestion will preserve and pricing will be inefficient. To this end, we redistribute a fraction of the toll revenue such that it serves as an incentive for higher PT usage. Let us denote a pricing scheme without any toll redistribution, "pricing". Meanwhile, the collected total toll can be distributed in two possible ways: (i) to expand the number of bus lines in operation and bus network coverage, so that the accessibility cost to bus (BA cost), decreases. We denote this strategy "pricing BA"; and (ii) in addition to (i), to provide a one-time incentive money for travelers who shift their mode from car to bus. We denote this strategy "pricing BAI" ("I" for

incentive). Both strategies can be readily implemented in the simulator or in a city where travelers utilize a public transport card, e.g. as in London. Regarding equity, thanks to the extension on VOT heterogeneity as introduced in Section 2.1, we investigate a modified version of strategy "pricing BAI", where the award money is only given to users who have low value-of-time (presumably those who have lower income). Denote this strategy as "pricing BAIp" ("p" stands for group "poor"). We expect that the design and operation of such incentive programs in practice are also feasible, however highly dependent on political, financial and other influences. We do not wish to deepen our discussion into the direction of policy-making, but to demonstrate via compact analysis that proper distribution of toll revenue for PT systems can have critical impact on the efficiency and equity of congestion pricing. We will show in the next section the resultant performance of each pricing strategy.

3. System performance under the pricing strategies

In this section, we demonstrate the performance of the proposed pricing strategies with a case study from the agent-based simulation. The set-up of the case study is given in Section 3.1. In Section 3.2, system performances of the different pricing strategies are compared and discussed. We now describe the chosen performance indicators. The primary goal of pricing is to alleviate congestion in road networks. Thus, space-mean traffic density and the total travel time of all travelers (person-hours traveled, PHT) are estimated as the measurement of traffic performance. Secondly we focus on the impact of pricing on multimodal mobility and the efficiency of the pricing strategies. The efficiency is defined as the ratio between the total saving of PHT (sPHT) and the total toll paid (TTP) by all travelers. To have the same unit of the two items, TTP is converted to time units by VOT, denoted as vTTP. The impact on multimodality is considered as the third measure, which is basically represented by the change of mode share of bus. Furthermore, we examine in Section 3.3 how user's behavioral response to the pricing over time and how the pricing rates are updated accordingly. These results will illustrate the mechanism and convergence feature of the proposed pricing scheme. In the last sub-section, we present a sensitivity analysis on the control parameters of the proposed pricing controller as in Eq. (2).

3.1. Case study set-up

The case study is carried out in the well-known Sioux-Falls urban network. The structure of the network captures the major arterial roads of the real city: highways with 3 lanes per direction at the perimeter of the network and urban roads with 2 lanes in the city center. The average link length is 0.4 km. The total length of the network is 150 lane-kms. The scale of the network is suitable for applying MFD-based analysis. The embedded public transport system serves five bi-directional bus lines crossing the network (in total ten different operation bus routes). The spacing between bus stops is 0.6 km and the frequency of service ranges from 5 to 15 min. The studied network area holds a total amount of 110,000 travelers. The daily plans of the travelers are created based on detailed census data. The generated traffic demand exhibits typical morning- and evening-peak characteristics. For more information on the design of the Sioux-Falls simulation scenario, readers may refer to Chakirov and Fourie (2014). An area-based pricing is implemented for the whole network. Agents need to pay for the toll once they decide to travel by car in the network independently of the trip length.

The steps presented in Section 2.2 are described in more details for the specific case study. The first task is to investigate the existence of a well-defined and low-scattered network level MFD and then identify the critical density K_{cr}. As the MFD depends on the property of a network, one can obtain parameter K_{cr} from historical data, e.g. analyzing the speed-density relationship of the network for a few typical weeks. In our case, K_{cr} is estimated based on preliminary simulation results when pricing is not implemented. Given that the traffic model of the agent-based simulator is simplistic and the route choice adaptation is not well-captured, the network flow contains quite some scatter. Nevertheless, if the network speed-density relation is well represented, it can be utilized to estimate the K_{cr} value. Speed and density data are collected from scenarios of different demand profiles (population size, origin–destination loading, etc.), and the speed-density Macroscopic Fundamental Diagram is displayed in Fig. 2(a). Each data point corresponds to a network-level space mean during a 5-min time interval. This speed-density curve is fitted by a natural-logarithm-family function and a best fit curve is obtained $v = 42 - 9.46 \log k$ (the value of the R-square test is 0.95). Alternative functions may also be applied to approximate the shape of the MFD. Given the fitted function and the dynamic equation $Q = kV$, where Q, K, V are network space mean flow, density and speed respectively, the critical density that maximizes network flow is obtained by the solution of the equation $\frac{dQ}{dK}\big|_{K=K_{cr}} = V(K_{cr}) + K_{cr}\frac{dV}{dK}\big|_{K=K_{cr}} = 0$. Fig. 2(b) displays this MFD. Note that the y-axis is outflow or the trip finish rate of the network, which is a scaled version of flow (related to the average trip length, see in Geroliminis and Daganzo, 2008; Leclercq et al., 2015). We observe that K_{cr} does not vary from iteration to iteration and its value is approximately 32 veh/km. Note that the network experiences many states with values larger than K_{cr}. Thus a time-dependent area-based pricing scheme is expected to decrease network delays.

User heterogeneity is treated numerically as follows. We define and implement two user groups in the case study: A group of users with a high value of VOT, which takes one third of the whole population, and a group with lower VOT for the rest of the population. We consider that the mean VOT for the entire population has a value of 16 \$/h. This value is derived from Axhausen et al. (2007) who studied the VOTs of different types of Swiss travelers. Given this value, we assign the VOT for the two groups to be 26 \$/h and 11 \$/h respectively.

3.2. Network traffic performance

The simulations are executed following the algorithmic steps described in Fig. 1. A value of 0.5 $/(veh/km) and 0.3 $/(veh/km) is used for c_1 and c_2 in Eq. (2). Let us first examine the resultant traffic performance of the network, without and with the implementation of pricing. Denote the two cases as "no pricing" and "pricing", correspondingly. The density time series of these two cases (left axes, the lines in blue and green) are displayed in Fig. 3(a). Time series of the final optimal pricing is plotted in the same figure, where *y*-axis values are shown on the right-hand-side of the figure. The final pricing gives a 0.8$/trip toll for the morning peak between 8 h and 8h30, and a higher step toll for the evening peak between 17 h and 19 h. We can observe that the congestion level of the network, as expressed by the density, decreases significantly for both morning and evening peaks, albeit under different toll rates (see the secondary *y*-axis of the same graph).

In the same figure, we also display density and toll series when "pricing BA" scheme is implemented. Recall that "pricing BA" improves the accessibility to PT service, e.g. the resultant average access (walking) time from home to bus stops is assumed to decrease by 15%. Comparison of the results of this strategy with the other two cases highlights the importance of the proposed congestion pricing scheme with a preferential treatment of PT. With the improvement in accessibility, we find that congestion level can be further reduced (comparing density time series). It is intriguing that this is achieved at an even smaller amount of toll. Estimated from the result, this strategy succeeds in maintaining network density under 40 vehs/km with 20% less toll charge. An interesting observation is that even if the set-point of the PI controller is set at a value of K_{cr} = 32 veh/km, the network operates at densities close to 40 veh/km. This is the value that the system converges after many iterations and not a higher price is chosen. The main reason is that the total cost of traveling by bus is about the same with the cost traveling by car plus the toll for this specific density range. Thus, even if a higher price would be chosen, this would result in worse system performance (higher total cost and unequal travel cost between the two modes) as public transport is not any more attractive. This result motivates us to take a step further and we will show later that when incentives are given to the users and PT is improved, the network will operate at lower car density values. Note that the optimal tolls of both pricing strategies change smoothly and do not oscillate over time, making them feasible to implement in practice. Note also that in Fig. 2(b) that the outflow of the network for density close to 40 veh/km is only 2.5% less than the maximum outflow of K_{cr}. If one also considers the scatter of the MFD around these values, the difference is negligible.

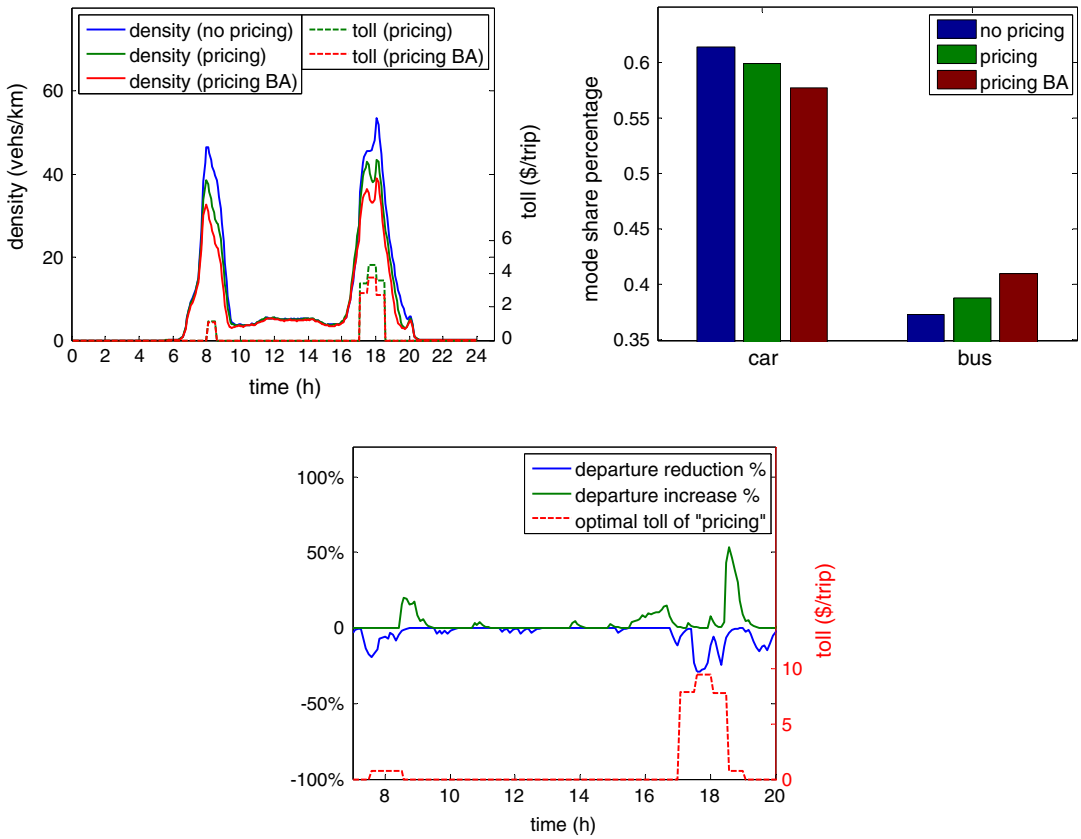

Fig. 3. Comparison of the system performance of "no pricing", "pricing" and "pricing BA": (a-top left) Density time series and the toll rates (dashed line); (b-top right) mode share; and (c-down) shift of departure time, without and with optimal time-dependent pricing (the dashed line: toll rates).

Table 1
Traffic and welfare gain among the three pricing strategies.

Strategies	PHT (h)	sPHT (h)	vTTP (h)	Efficiency ratio (%)
Pricing	69,124	3449	10,644	32.4
Pricing BA	64,165	8408	7626	110.2
Pricing BAI	64,785	7788	6602	118.0

Fig. 3(b) illustrates the mode share of buses under the three pricing strategies. As expected, the number of bus users increases with pricing implementation and improvement of level of service. Strategy "pricing BA" triggers a 5% increase in mode shift than "pricing", and a nearly 10% increase comparing to "no pricing". Fig. 3(c) displays the change of departure behavior, after pricing (the "pricing" scheme) is implemented. The time series in green[2] (positive values) represents the increase in departure during that time interval, comparing to the "no pricing" case, while the one in blue (negative values) reflects the reduction of demand who attempt to avoid the cost of toll. Note that the values of the two curves do not add to zero, because there is a significant amount of travelers prefer to change mode to remain traveling in the toll period without increasing their travel cost. Nevertheless, it is clearly shown that travelers who persist in using cars attempt to avoid paying the tolls. During the tolling period, reduction of departures can be observed. On the other hand, increases of departures are clearly shown before and after the tolling period, especially in the evening when congestion is more evident. When traveling during pricing period is necessary (e.g. during the evening peak 17–19 h), users choose to avoid paying higher tolls which is reasonable. Note that in many pricing-related studies, public transport is not considered as an option. Therefore the benefit of pricing comes from shift of departure time, while in this study bus serves as a competitive and attractive mode and mode shift plays an important role in congestion reduction. Another interesting observation is that roughly similar amount of car-usage reduction can be achieved in the morning by enforcing a much smaller toll rate than in the evening. To understand this result, we will have a careful look on how densities and the corresponding pricing rates change over simulations in Section 3.3.

To quantify the effectiveness of the pricing strategies, we calculate the improvement in travel time (the PHT) over toll adjustments, at an aggregated scale for cases "no pricing" and "pricing". Under congestion pricing, the total PHT decreases from 72,573 h to 69,124 h by 7%, while the average travel time by individual car travelers drops from 0.53 to 0.46 h/trip. There is however no change for traveling by bus. The proposed pricing scheme is able to avoid highly congested states. Nevertheless, travelers pay a high amount of toll to obtain such a decrease in PHT.

To quantify the efficiency of the pricing strategies, we estimate the sPHT (savings in travel time) and vTTP (the required toll paid) which is the extra cost to the users. Savings are the difference in PHT comparing to the "no pricing" case. The sum of these two terms is the "gain" of a pricing strategy. Traditionally from an economic point of view, pricing is considered a transfer of money from car travelers to system operators and is not integrated in the objectives of the design. While the scientific literature of pricing is vast, pricing implementations are very few and this might be because people are reluctant to pricing and acceptability is limited. A smart and sustainable pricing strategy would hopefully result positive gains, without imposing high expenses to the users. In Table 1, savings in travel time and efficiency ratio are estimated for the three pricing strategies: "pricing", "pricing BA", and "pricing BAI". The resultant PHT of the "no pricing" scenario is 72,573 h. The efficiency ratio is obtained by dividing sPHT and vTTP. Since a fraction of the collected toll is returned to users, the toll value vTTP should not be entirely viewed as cost term. Therefore from a total-cost perspective (i.e. PHT + vTTP), strategies "pricing BA" and "pricing BAI" are even more efficient as part of the tolls are redistributed internally among the users.

The resultant efficiency ratios shown in Table 1 indicate, evidently, that pricing can be remarkably efficient when the toll revenue is invested in improving accessibility of bus service, as well as when the mode shift to high-occupancy buses is subsidized. Strategy "pricing BA" and strategy "pricing BAI" are able to reach an efficiency value larger than 110%. Note that in the classical morning commute literature with a constant capacity bottleneck, pricing efficiency is at a maximum of 100%, i.e. the total delay savings are equal to the total toll paid. Nevertheless, in case of urban networks with MFD representations, the savings can be larger because the network operates at higher flows and congestion period is shorter. Note in Fig. 2(a) that the pricing interval is smaller than the interval of congestion in the no pricing case. Another interesting observation is that savings in travel time is smaller, when incentive is provided for mode shift (7788 h instead of 8408 h). The explanation is given as follows. A significant amount of users switch to buses, given that they value the award money for doing so. However, traveling with bus is not as fast as with cars, since congestion does not exist anymore in the car network. As a result, Strategy "pricing BAI" could not further improve the total travel time. Nevertheless, this strategy requires less toll to satisfy the congestion management goal (because users are more willing to shift from car to bus), showing the highest efficiency and equity, which will be further discussed in the next sub-section.

3.3. User adaptation and long-term stabilization

The objective of MFD-based traffic management strategies including the one developed in the paper are to monitor network density and try to retain it at a desired level that maximizes system outflow, without the direct need for demand

[2] For interpretation of color in Fig. 3, the reader is referred to the web version of this article.

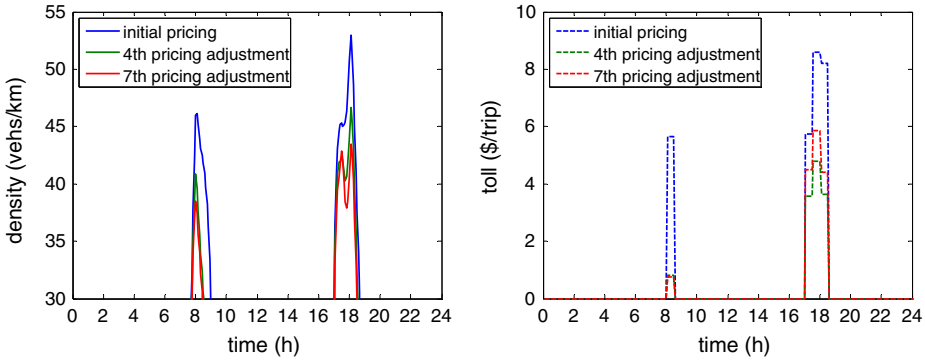

Fig. 4. (a-left) density series and (b-right) toll rates of three iterations before reaching the optimal "pricing".

information or prediction. We have mentioned several of the state-of-the-art traffic signal control approaches in Section 2.2, which are based on the MFD. In the current work, we intend to apply the same principle by utilizing a time-dependent area-based pricing, which creates additional difficulties and complexities. Firstly, pricing influences travelers' choices in non-fully predictable ways, whereas in the ramp-metering or network-level perimeter control car users have to obey the controlled traffic signals. Secondly, traffic signals mainly influence the delay of drivers after they make a decision to travel by car and adapt signal settings in an online framework, while pricing attempts to change the behavior of a traveler before she starts her trip (by influencing time of departure or mode choice). This is a long-term process, as the traveler has to learn how to react in these changes. The controller that determines the dynamic pricing for influencing mode choice, considers both the network congestion, represented by the different between density k and critical density K_{cr} (2nd term RHS of Eq. (2)), and user's behavioral change, represented by the evolution of density over pricing adjustments (3rd term RHS of (2)), simultaneously. This is innovative, given that such control scheme is rarely reported in the literature of pricing. We have demonstrated that such control approaches can also be successfully applied with congestion pricing. An interesting question to investigate then is how users adapt their travel choices in short- and long-term periods, and how pricing responds interactively with those changes.

We now investigate the evolution of network congestion densities, toll rates, total toll paid (TTP) and the total number of mode shifters (from using car to bus) over toll adjustments. We also look into how travelers are adapting between the toll adjustments. As mentioned before, we do not intend to change the toll in a day basis, as we need to provide time until the system reaches some equilibrium. Fig. 4(a) and (b) display the time series of density and pricing of three toll adjustments, $k = 1, 4, 7$, under strategy "pricing" before reaching the final optimal toll (in total 8 adjustments were performed). This graph shows that the initial tolls ($k = 1$) are almost equally high for both morning and evening peak periods. The morning traffic adapts quickly to the increased cost due to pricing, and the density level decreases to slightly above K_{cr}. The morning toll is then lowered so that car users are not overcharged (see the graph for $k = 4$). It can be observed here that pricing decreases with density. Nevertheless as density $K_k(t)$ is higher than K_{cr}, it makes the term $c_1(K_k(t) - K_{cr})$ (2nd term RHS of Eq. (2)) positive. Then a careful calibration of the c_2 parameter leads to this result. If we would apply a simpler controller (such as the one in Zheng et al. (2012)), this trend is not feasible. In other words, if $c_2 = 0$, the pricing is not efficient.

We can observe that there is almost no induced demand for the morning peak, even though the toll rates become three times less comparing to the initial toll. While for the evening peak, it seems that higher toll must be imposed and kept. The reason is twofold: (i) travelers have to travel during the peak hour to perform their activities without scheduling penalties in their utility. Leaving activities early would experience "schedule penalty" which is larger than the cost of pricing and (ii) the PT mode bus is not sufficiently fast to trigger mode shift until the cost of pricing is greater than the difference in travel times between car and bus. Therefore many users choose to travel even after pricing is applied, which weakens the elasticity of demand towards pricing.

A third possible explanation of demand inelasticity lies on the fact that users who choose to travel with cars in the morning, have to travel with cars in the evening (e.g. to not leave their cars at work places), which constraints their mode choice in the evening and forces them to take the high congestion tolls. However, as users (agents) can learn their experienced cost iteratively, they have the opportunities to choose the other alternative mode (bus) and keep this plan if such change results a higher utility by the end. Please note that the travelers if they decide to travel by car, they need to return the car at home by the end of the trip, which constrains their alternatives, for example it is not possible to drive to work in the morning and return by bus due to a high evening toll. Thus we believe the high pricing cost should have sufficient impact on mode choice. However, when users finish their activities in the evening, they are forced to leave their work place since there is no utility of staying late. Then, many car users travel anyway despite the high toll, which explains the cause of the inelasticity to pricing. An alternative would be to travel by bus for all their trips during the day, as it is not possible to leave the car at work and return by bus. Nevertheless pricing adjustment to user's adaptation can be observed for the evening peak as well. As the implementation of the initial toll decreases network density, a considerable drop of the toll rate can be found for $k = 4$ in Fig. 5(b). Given this adjustment, it seems that travelers begin to switch back to the peak period and take advantage of

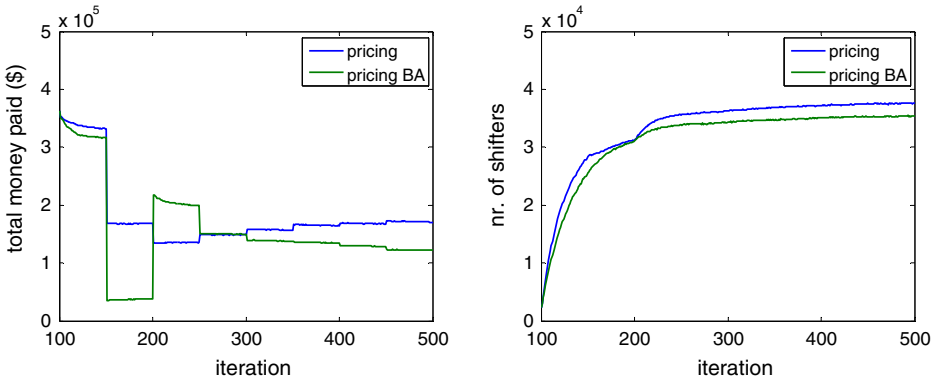

Fig. 5. Evolution of (a-left) total money paid over iterations, and (b-right) the amount of mode shifters over iterations under strategies "pricing" and "pricing BA".

the improved network condition. Our pricing scheme identifies this trend. The same figure shows that a higher toll is re-implemented for $k = 7$. Recall that a pricing adjustment takes place approximately every 50-iterations of the agent-based simulation (after 50 iterations user behavior is stable, and the behavioral change after more iteration is negligible).

A relevant question now is how sensitive are the travelers to the change of tolls and how fast they adapt. Fig. 5(a) plots TTP for strategies "pricing" and "pricing BA" over the entire iterations. The initial pricing is implemented at the 100th iteration. Since then, the tolls are adjusted every 50 iterations. It can be seen that travelers do not have previous experience and overreact when the tolls are initially introduced, reflected by the huge jumps at the 150th ($k = 2$) and the 200th ($k = 3$) iterations. Similar phenomenon can be observed in Fig. 5(b), where the number of mode shifters (from car to bus) is plotted over iterations for the same two cases. After the 6th adjustment of the toll (the 350th iteration), both TTP and mode shifts change slowly over toll adjustments. With sufficient amount of iterations, it is expected that user-equilibrium is achieved. These are promising results, as they indicate that the developed pricing scheme shows smooth convergence in the long-term operation that involves user adaptation. It is expected that a small level of induced demand might appear from time to time, and the system would be able to adjust and avoid congestion without performing drastic behavioral changes. Note also that the total money paid is significantly smaller for the final pricing (after iteration 500 or $k = 7$), which highlights the importance of pricing adaptation after the first implementation. The efficiency ratio of pricing for $k = 1$ is less than 10%.

Let us now compare the price controller (2) with the controller developed in Zheng et al. (2012) for determining a peak-hour cordon pricing scheme. Zheng et al. (2012) carried out the first work that introduces the concept of MFD for the update of congestion prices, where the authors employed an I-type of controller optimizing prices to keep the network operate at non-congested states. However, the authors only determined single price rates for the morning and the evening peak hours respectively, which lack flexibility in dealing with fluctuation of congestion states within each peak-hour. It also lacks flexibility in adapting the prices, since (I) public transport mode was not modeled sufficiently and provided as an alternative mode, and (II) prices are updated without considering how congestion evolves over price adjustments. In this paper, limitation (I) above is treated in our simulation model and pricing strategies treating public transport are discussed, while (II) is incorporated by the third item at the RHS of the regulator in Eq. (2). Furthermore, the proposed pricing scheme shows an improved pricing efficiency. We executed a few simulation scenarios of different random seeds, applying the controller of Zheng et al. (2012) and the proposed "pricing" strategy of controller (2) without treatment on improving bus service or distributing mode shift incentives.

The resultant efficiency ratio is found on average 23% which is nearly 10% less than controller (2) which has a mean value of 32%. It is not surprising that such improvement can be achieved, given that user adaptation is taken into account in the pricing update process, which helps avoid overpricing on the users. As parameter c_2 in Eq. (2) does not exist and the bus performance is worse than the cost of car travelers at K_{cr}, the pricing mechanism will continue to increase the pricing despite the fact that mode shift is constraint due to bus performance and results in overpricing.

3.4. Sensitivity analysis on the control parameters

Given the complexity of the system dynamics, analytical derivation of the two parameters c_1 and c_2 in price controller (2) is extremely difficult, as these dynamics are influenced by both user adaptation and mode choice. Therefore parameters c_1 and c_2 were determined via trial-and-error. In this sub-section, our objective is to show through a sensitivity analysis that system performance does not differ significantly when proper values of the parameters are chosen, and the proposed pricing controller shows robustness in congestion reduction. To this end, we vary the values of both c_1 and c_2 and execute the simulations under five combinations of the parameters. This variation is represented by the percentages that the values of the two parameters are changed, comparing to the values utilized in a base scenario. Due to the computational cost of the

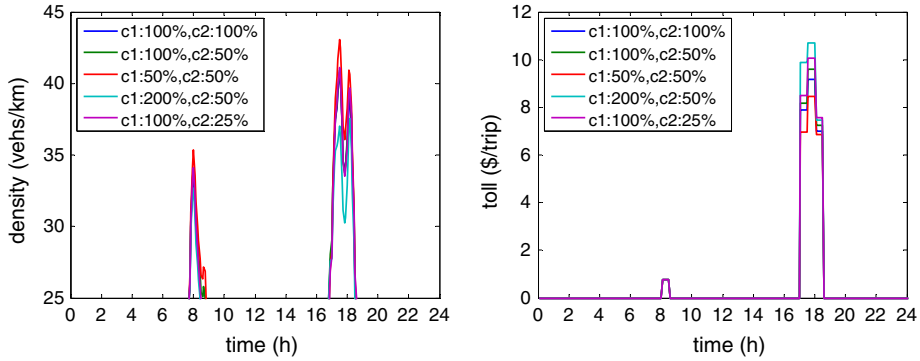

Fig. 6. The resultant (a-left) density series and (b-right) toll rates of five scenarios employing different values of parameters c_1 and c_2.

simulation runs, we investigate and present the results of five combinations, while a more comprehensive analysis would be tedious and similar conclusions can be drawn.

All the five scenarios reach the agent-based equilibrium where the average cost of the users nearly keeps unchanged. Fig. 6 displays the resultant density series and toll rates under equilibrium. The different values given to c_1 and c_2 are listed in the legend (using this metric, the base scenario owns c_1 100% and c_2 100%). According to the figure, it is evident that the values of the parameters affect the final congestion level and optimal toll. Let us compare these two performance indicators between an aggressive toll scheme where price changes with higher magnitude (for example in the case of c_1 200%, c_2 50%) and a mild toll scheme (the case of c_1 50%, c_2 50%). The aggressive scheme imposes higher toll rates, which consequently lead to a larger reduction of congestion (lower density level on average). Now if we compare the efficiency ratio of the two toll schemes to the one of the base scenario, we find that the aggressive toll scheme gives almost the same ratio (32.3%) as the base scenario (32.4%) while the mild scheme shows a slightly smaller ratio (25.6%). The total travel time (PHT) for the three cases are 68,413, 69,124 and 69,988 respectively.

As a general remark, the system performance appears not very sensitive to the fluctuation of the chosen values of the control parameters. While further case studies might be needed, we argue that the efficiency of the final pricing is not significantly dependent on the choice of parameters if a range of proper values can be found in advance via trial-and-error. It should be noted that differences in the final system performances can still be identified, which are as expected. Nevertheless, it is infeasible to obtain optimal control parameters via analytical ways using classical control theory approaches, for transportation systems of complex behavioral dynamics.

4. The impact of user heterogeneity

The last section of this paper is dedicated specifically to the impact of pricing on users with different VOTs. We consider that the heterogeneity in VOT, which are obtainable from surveys, causes differences in mode choice behavior and the perceived travel costs. Utilizing this setting, we investigate the responses of two user groups to congestion pricing, and the equity conditions of various pricing strategies. We aim to propose feasible and equitable pricing strategies where the savings are in favor of the more vulnerable groups.

The simulation scenarios (in terms of simulation setting, random seeds, etc.) remain the same as presented in the previous section. Only that now these scenarios are extended with two groups of users having two different VOTs, and we analyze the results more at a disaggregated user-group level.

4.1. Heterogeneity in mode choice

Let us start with the observations on the behavioral difference of users, when heterogeneity is introduced. We demonstrate this in the "no pricing" case. A comparison of the resultant mode share of buses between the two groups is displayed in Fig. 7(a) for scenario without user heterogeneity (a single user group, thus the mode split has the same value) and the scenario with the two groups defined earlier. Denote the user group with high VOT the "R group" and the one with low VOT the "P group" (as rich vs. poor groups). It is observed that the two groups have strong contrast in mode choice behavior. While the mode share of buses is identical for the single-group scenario, there is nearly 10% difference between the R group and the P group when two VOTs are differentiated. It is obvious that the R group tends to prefer traveling with cars, as it is faster. This is logical because travel time is valued much higher for the users of this group than those of the P group. While for the P group, mode share of buses increases even if the utility difference remains the same between traveling with the two modes. The reasoning is exactly the opposite of the one for the R group: Since now the P group has a smaller VOT, travelers care less about the cost of travel time therefore are more "willing" to choose traveling by bus.

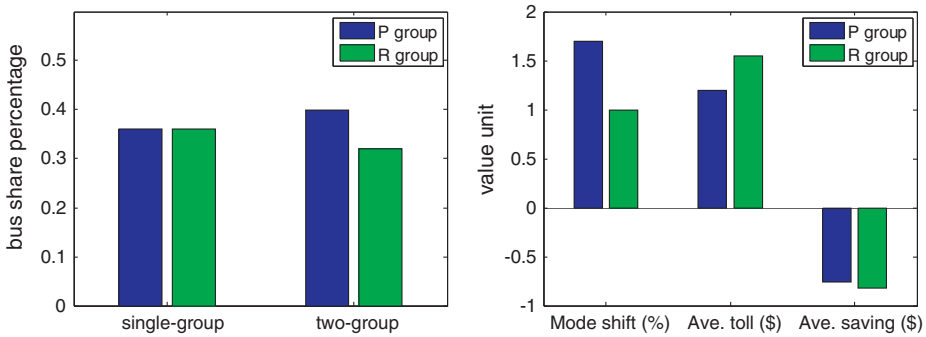

Fig. 7. (a-left) Mode choice difference of the two user groups, without congestion pricing, and (b-right) comparison of mode shift, toll paid and the net savings between the two user groups under strategy "pricing".

When congestion pricing is applied, the two user groups exhibit distinct reactions which are as expected. Fig. 7(b) displays the comparison of the resultant mode shift (the percentage of users shift from cars to buses after pricing), average toll paid per person (total toll paid divided by the total amount of the paying users), and the average net savings per person (net saving is the difference between the reduction in total travel time and the total toll paid by all users) under strategy "pricing". It is found that more users of the P group tend to switch to buses, though the numerical results seem to suggest a minor difference between the two groups. The resultant average pricing rates (toll) indicate that the R group users are willing to pay 0.3$ more for traveling faster with cars. The net saving is negative for the entire population, and almost equivalent for the two groups. These results are consistent with the findings summarized in Levinson (2010) which highlighted the difference in behavior of travelers with different VOTs. We will show next that pricing strategies that treat in a preferential way public transport, can have positive savings for the users (delay savings are higher than toll paid).

4.2. Distributional effect under different pricing strategies

We now aim to shed some light in the equity perspective of the proposed pricing strategies. We define here equity as the (average) savings in travel cost which include both travel time and toll paid. Fig. 8(a) illustrates the mode shift by strategies "pricing", "pricing BA" and "pricing BAIp" (recall that strategy "pricing BAIp" distributes subsidy only to the mode shifters from the "P group"). Mode shift here calculates the mode shift to bus under each of the three strategies compared to the "no pricing" case. It can be seen that as the incentive of using buses arises, there is an increasing demand for both groups to travel by bus. This is not only because pricing increases the cost of car, but also because PT benefit increases the attraction of bus. For the R group, it is evident that users are reluctant to change mode even though traveling with bus becomes less costly. Mode shift is found below 1% under strategy "pricing" while less than 2.5% under the other two pricing strategies. This is probably due to the fact that traveling with cars becomes less costly at the same time as more users travel with buses, making travel by car remain attractive for the R group that values travel time more importantly. While for the P group, the resultant mode shift shows a clear tendency of growth as the strategies provide more benefit by improving the bus level of service. Under strategies "pricing" and "pricing BA", the mode shifts are 1.8% and 4.5% respectively, almost double for the mode shift of the R group. Under "pricing BAIp" the percentage goes to 6.8% which nearly triples that of the P group. Comparing to strategy "pricing BA", "pricing BAIp" makes 3% more users of the P group switch their mode.

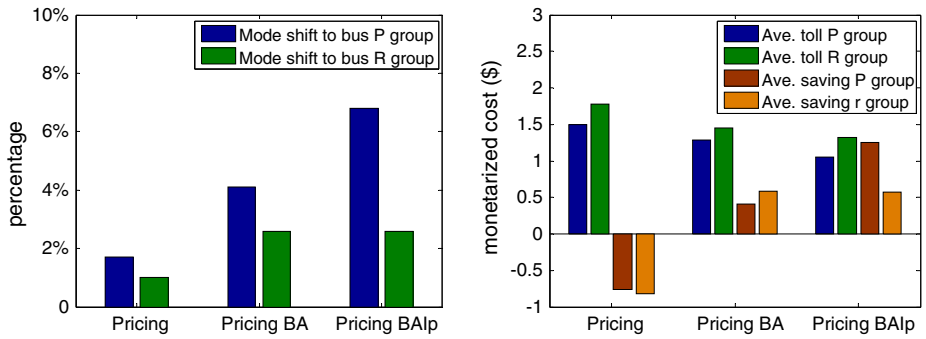

Fig. 8. (a-left) Comparison of the difference in mode shifts and (b-right) comparison of mode shift, toll paid and savings the two user groups among the three pricing strategies: pricing, pricing BA and pricing BAIp.

Fig. 8(b) illustrates the comparison of average toll paid and average savings for the two user groups under the three pricing strategies. In all cases, the R group users are willing to pay higher toll to reserve the right of traveling with cars. Note that both user groups benefit from the PT incentives, since they all get positive savings. The savings of the P group users outweigh the R group by 100% when subsidy only provided to the R group. As the users of the P group are generally considered to represent the users with lower income (Börjesson and Eliasson, 2014), strategy "BAIp" is undoubtedly a more equitable pricing scheme. These results indicate that heterogeneity of users and distributional effect should be treated carefully when designing strategies and policies of congestion pricing, so that beneficial and equitable conditions can be achieved for all travelers.

5. Discussion

In this paper, we developed and investigated time dependent area based pricing strategies for congested multimodal urban networks with consideration of two user groups with heterogeneous value-of-time. A time-dependent area-based pricing scheme was proposed, where the tolls are iteratively adjusted through a Proportional–Integral type feedback controller, based on both the level of congestion represented by the MFD, and user's adaptation to the toll cost. With this pricing scheme, we integrated incentive schemes to improve PT accessibility and services for motivating mode shift and increasing the efficiency of the pricing and investigated the impact of pricing on the two groups of heterogeneous users with different values-of-time. The pricing strategies were tested in an agent-based traffic simulation model, to ensure that behavioral changes by pricing are well captured. Case study of the Sioux Fall network revealed appealing results: (i) the proposed toll scheme is highly effective in congestion reduction, (ii) comparing to similar toll schemes in literature the proposed pricing scheme is more flexible in toll adjustment as it takes user adaptation into account, and showed a smooth user equilibrium in long-term operation (iii) significant difference in behavioral responses and trip costs were found for the two groups of users, (iv) by integrating incentive programs, the pricing scheme achieved large welfare gain for all user groups and (v) by realizing user heterogeneity, pricing strategies can be designed to achieve more equitable result. The output of this paper can be of great importance to the existing pricing-based management strategies in practice, for example to support the optimization of the congestion toll scheme in different cities worldwide. On-going work further investigates the equity of the pricing schemes, e.g. welfare distribution among different trip purposes and impact on trip departures. How such pricing scheme can help maintain network performance under network disruption scenarios such as road construction (encouraging detour around the bottlenecks) will also be under investigation.

Acknowledgements

This research was financially supported by ERC Starting Grant "METAFREW": Modeling and controlling traffic congestion and propagation in large-scale urban multimodal networks. The authors would like to thank Dr. Rashid A. Waraich and Prof. Kay Axhausen from Institute for Transport Planning and Systems, Swiss Federal Institute of Technology Zurich, for his assistantship with MATSim and for providing the Sioux Falls network.

References

Aboudolas, K., Geroliminis, N., 2013. Perimeter and boundary flow control in multi-reservoir heterogeneous networks. Transp. Res. Part B 55, 265–281.
Ampountolas, K., Zheng, N. and Geroliminis, N., 2014. Perimeter flow control in bi-modal urban road networks: a robust feedback control approach. In: Proceedings of the 13th European Control Conference, June 24–27. Strasbourg, France, pp. 2569–2574.
Arnott, R., 2007. Congestion tolling with agglomeration externalities. J. Urban Econ. 62 (2), 187–203.
Axhausen, K., Hess, S., Koenig, A., Bates, J., Bierlaire, M. and Abay, G., 2007. State-of-the-art estimates of Swiss value of travel time savings. In: Paper Presented at the 86th Annual Meeting of Transportation Research Board. Washington, DC, USA.
Börjesson, M., Eliasson, J., 2014. Experience from the Swedish value of time study. Transp. Res. Part A 59, 144–158.
Börjesson, M., Cherchi, E., Bierlaire, M., 2014. Within-individual variation in preferences: equity effects of congestion charges. Transp. Res. Rec. 2382, 92–101.
Chakirov, A. and Fourie, P., 2014. Enriched Sioux Falls Scenario with Dynamic and Disaggregate Demand, Working paper. Future Cities Laboratory, Singapore ETH Centre, Singapore.
Chiabaut, N., 2015. Evaluation of a multimodal urban arterial: the passenger macroscopic fundamental diagram. Transport. Res. Part B 81, 410–420.
Chiabaut, N., Xie, X., Leclercq, L., 2014. Performance analysis for different designs of a multimodal urban arterial. Transportmetrica B 2 (3), 229–245.
Daganzo, C., 2007. Urban gridlock: macroscopic modeling and mitigation approaches. Transp. Res. Part B 41 (1), 49–62.
De Palma, A., Lindsey, R., 2011. Traffic congestion pricing methodologies and technologies. Transp. Res. Part C 19 (6), 1377–1399.
Eliasson, J., Mattsson, L., 2006. Equity effects of congestion pricing: quantitative methodology and a case study for Stockholm. Transp. Res. Part A 40 (7), 602–620.
Geroliminis, N., Boyacı, B., 2012. The effect of variability of urban systems characteristics in the network capacity. Transp. Res. Part B 46 (10), 1607–1623.
Geroliminis, N., Daganzo, C., 2008. Existence of urban-scale macroscopic fundamental diagrams: some experimental findings. Transp. Res. Part B 42 (9), 759–770.
Geroliminis, N., Levinson, D., 2009. Cordon pricing consistent with the physics of overcrowding. Transport. Traffic Theory, 219–240.
Geroliminis, N., Zheng, N., Ampountolas, K., 2014. A three-dimensional fundamental diagram for mixed bi-modal urban networks. Transp. Res. Part C 42, 168–181.
Godfrey, J., 1969. The mechanism of a road network. Traffic Eng. Control 11 (7), 323–327.
Gonzales, E., Daganzo, C., 2012. Morning commute with competing modes and distributed demand: user equilibrium, system optimum, and pricing. Transp. Res. Part B 46 (10), 1519–1534.
Gonzales, E., Daganzo, C., 2013. The evening commute with cars and transit: duality results and user equilibrium for the combined morning and evening peaks. Transp. Res. Part B 57, 286–299.
Haddad, J., 2015. Robust constrained control of uncertain macroscopic fundamental diagram networks. Transp. Res. Part C 59, 323–339.

Haddad, J., Ramezani, M., Geroliminis, N., 2013. Cooperative traffic control of a mixed network with two urban regions and a freeway. Transp. Res. Part B 54, 17–36.
Ji, Y., Geroliminis, N., 2012. On the spatial partitioning of urban transportation networks. Transp. Res. Part B 46 (10), 1639–1656.
Ji, Y., Luo, J., Geroliminis, N., 2015. Empirical observations of congestion propagation and dynamic partitioning with probe data for large-scale systems. Transp. Res. Rec. 2422 (2), 1–11.
Keyvan-Ekbatani, M., Yildirimoglu, M., Geroliminis, N., Papageorgiou, M., 2015. Multiple concentric gating traffic control in large-scale urban networks. IEEE Trans. Intell. Transp. Syst. 16 (4), 2141–2154.
Lawphongpanich, Yin, Y., 2010. Solving the pareto-improving toll problem via manifold suboptimization. Transp. Res. Part C 18 (2), 234–246.
Leclercq, L., Geroliminis, N., 2013. Estimating MFDs in simple networks with route choice. Transp. Res. Part B 57, 468–484.
Leclercq, L., Chiabaut, N., Trinquier, B., 2014. Macroscopic fundamental diagrams: a cross-comparison of estimation methods. Transp. Res. Part B 62, 1–12.
Leclercq, L., Parzani, C., Knoop, V., Amourette, J., Hoogendoorn, S., 2015. Macroscopic traffic dynamics with heterogeneous route patterns. Transp. Res. Part C 59, 292–307.
Levinson, D., 2010. Equity effects of road pricing: a review. Transp. Rev. 30 (1), 33–57.
Liu, Z., Meng, Q., Wang, S., 2013. Speed-based toll design for cordon-based congestion pricing scheme. Transp. Res. Part C 31, 83–98.
Lou, Y., Yin, Y., Lawphongpanich, S., 2010. Robust congestion pricing under boundedly rational user equilibrium. Transp. Res. Part B 44, 15–28.
Lou, Y., Yin, Y., Laval, J., 2011. Optimal dynamic pricing strategies for high-occupancy/toll lanes. Transp. Res. Part C 19 (1), 64–74.
Lu, C., Mahmassani, H., Zhou, X., 2008. A bi-criterion dynamic user equilibrium traffic assignment model and solution algorithm for evaluating road pricing strategies. Transp. Res. Part C 16 (4), 371–389.
Mahmassani, H., Williams, J., Herman, R., 1987. Performance of urban traffic networks. In: Proceedings of the 10th International Symposium on Transportation and Traffic Theory. Amsterdam, The Netherlands, pp. 1–20.
Maruyama, T., Sumalee, A., 2007. Efficiency and equity comparison of cordon- and cordon-based road pricing schemes using a trip-chain equilibrium model. Transport. Res. Part A 41 (7), 655–671.
Meister, K., Balmer, M., Ciari, F., Horni, A., Rieser, M., Waraich, R., Axhausen, K., 2010. Large-scale agent-based travel demand optimization applied to Switzerland, including mode choice. In: Paper Presented at the 12th World Conference on Transportation Research. Lisbon, Portugal.
Nagel, K., Flötteröd, G., 2009. Agent-based traffic assignment: going from trips to behavioral travelers. In: Pendyala, R.M., Bhat, C. (Eds.), Travel Behaviour Research in an Evolving World. Emerald Group Publishing, pp. 261–293 (2012).
Nie, Y., Liu, Y., 2010. Existence of self-financing and pareto-improving congestion pricing: impact of value of time distribution. Transp. Res. Part A 44 (1), 39–51.
Ortigosa, J., Menendez, M., Tapia, H., 2014. Study on the number and location of measurement points for an MFD perimeter control scheme: a case study of Zurich. EURO J. Transport. Logist. 3 (3–4), 245–266.
Papageorgiou, M., Kotsialos, A., 2010. Freeway ramp metering: an overview. In: Proceedings of the 13th IEEE Conference on Intelligent Transportation System, September 19–22. Funchal, Portugal, pp. 228–239.
Pierce, G., Shoup, D., 2013. An evaluation of pricing parking by demand in San Francisco. J. Am. Plan. Assoc. 79 (1), 67–81.
Qian, Z., Zhang, H., 2013. The morning commute problem with heterogeneous travelers: the case of continuously distributed parameters. Transportmetrica A 9 (2), 178–203.
Ramezani, M., Haddad, J., Geroliminis, N., 2015. Dynamics of heterogeneity in urban networks: aggregated traffic modeling and hierarchical control. Transp. Res. Part B 74, 1–19.
Saeedmanesh, M., Geroliminis, N., 2015. Optimization-based clustering of traffic networks using distinct local components. In: 18th IEEE Conference on Intelligent Transportation System (ITSC). Gran Canary, Spain (in press).
Simoni, M., Pel, A., Waraich, R., Hoogendoorn, S., 2015. Marginal cost congestion pricing based on the network fundamental diagram. Transp. Res. Part C 56, 221–238.
Tian, L., Yang, H., Huang, H., 2013. Tradable credit schemes for managing bottleneck congestion and modal split with heterogeneous users. Transp. Res. Part E 54, 1–13.
Tsekeris, T., Geroliminis, N., 2013. City size, network structure and traffic congestion. J. Urban Econ. 76, 1–14.
van den Berg, V., Verhoef, E., 2011. Congestion tolling in the bottleneck model with heterogeneous values of time. Transp. Res. Part B 45, 60–78.
Verhoef, E., 2002. Second-best congestion pricing in general networks: heuristic algorithms for finding second-best optimal toll levels and toll points. Transport. Res. Part B 36 (8), 707–729.
Vickrey, W., 1969. Congestion theory and transport investment. Am. Econ. Rev. 59 (2), 251–260.
Wu, D., Yin, Y., Lawphongpanich, S., 2011. Pareto-improving congestion pricing on multimodal transportation networks. Eur. J. Oper. Res. 210 (3), 660–669.
Wu, D., Yin, Y., Lawphongpanich, S., Yang, H., 2012. Design of more equitable congestion pricing and tradable credit scheme for multimodal transportation networks. Transp. Res. Part B 46, 1273–1287.
Xiao, F., Qian, Z., Zhang, H., 2013. Managing bottleneck congestion with tradable credits. Transp. Res. Part B 56, 1–14.
Yang, H., Huang, H., 2005. Mathematical and Economic Theory of Road Pricing. Elsevier Science Inc., New York.
Yang, H., Zhang, X., 2002. Multiclass network toll design problem with social and spatial equity constrains. J. Transport. Eng. 128 (5), 420–428.
Yildirimoglu, M., Geroliminis, N., 2014. Approximating dynamic equilibrium conditions with macroscopic fundamental diagrams. Transp. Res. Part B 70, 186–200.
Yildirimoglu, M., Ramezani, M., Geroliminis, N., 2015. Equilibrium analysis and route guidance in large-scale networks with MFD dynamics. Transp. Res. Part C 59, 404–420.
Yin, Y., Lou, Y., 2009. Dynamic tolling strategies for managed lanes. J. Transport. Eng. 135, 45–52.
Yin, Y., Yang, H., 2004. Optimal tolls with a multiclass, bicriteria traffic network equilibrium. Transp. Res. Rec. 1882, 45–52.
Zhang, L., Levinson, D., Zhu, J., 2008. Agent-based model of price competition, capacity choice and product differentiation on congested networks. J. Transport Econ. Policy 42 (3), 435–461.
Zheng, N., Geroliminis, N., 2013. On the distribution of urban road space for multimodal congested networks. Transp. Res. Part B 57, 326–341.
Zheng, N., Waraich, R., Axhausen, K., Geroliminis, N., 2012. A dynamic cordon pricing scheme combining the macroscopic fundamental diagram and an agent-based traffic model. Transp. Res. Part A 46 (8), 1291–1303.
Zhu, Z., Du, L., Zhang, L., 2013. Rationing and pricing strategies for congestion mitigation: behavioral theory, econometric model, and application in Beijing. Transp. Res. Part B 57, 210–224.