Congestion pricing in a real-world oriented agent-based simulation context

Ihab Kaddoura*, Kai Nagel

Technische Universität Berlin
Transport Systems Planning and Transport Telematics
Salzufer 17–19; 10587 Berlin; Germany

* Corresponding author
E-mail: kaddoura@vsp.tu-berlin.de
Tel.: +49 30 314 78 793
Fax.: +49 30 314 26 269

December 21, 2018

Abstract

This paper investigates optimal congestion pricing strategies using a real-world oriented agent-based simulation framework which allows for complex user behavior. The applied simulation approach accounts for iteratively learning transport users, stochastic demand, and only approximates the user equilibrium, which may be considered as closer to real-world than a model where transport users behave completely rational, have a perfect knowledge about all travel alternatives, and travel behavior strictly follows the user equilibrium. Two congestion pricing rules are developed and investigated. The first one directly builds on the Pigouvian taxation principle and computes marginal external congestion costs based on the queuing dynamics at the bottleneck links; resulting toll payments differ from agent to agent depending on the position in the queue (QCP approach). The second one uses control-theoretical elements to adjust toll levels depending on the congestion level in order to reduce or eliminate traffic congestion; resulting toll payments are the same for all travelers per time bin and road segment (LP approach). The pricing rules are applied to Vickrey's bottleneck model and the case study of the Greater Berlin area. The simulation experiments reveal that with and without mode and departure time choice, the rather simple LP rule results in a higher system welfare compared to the more complex QCP approach. The LP rule appears to better take into account the system's dynamics and the agents' learning behavior. The results also reveal that pricing significantly reduces traffic congestion, however, there is still a remaining delay, even with departure time choice. Overall, this paper points out further need for research and contributes to the exploration of optimization heuristics for real-world oriented simulation approaches.

Keywords: Congestion pricing; Marginal social cost pricing; Pigouvian taxation; Agent-based simulation, optimization heuristics

1 Introduction and problem statement

Following the concept of marginal social cost pricing, introduced by Pigou (1920), the social welfare is maximized by setting tolls equivalent to marginal external costs. Hence, optimal toll setting requires knowledge about the marginal congestion cost which in turn requires knowledge about the congestion function and demand elasticity.

A possible way to model road traffic congestion is the application of a so-called capacity restraint function (CR-function) to translate the traffic volume and road-specific parameters into a travel time. Based on this approach, all travelers on the same road segment are assumed to experience the same travel cost. The above works for static situations. For dynamic situations, a possible way is to apply a queue model which accounts for dynamic traffic congestion, i.e. a queue that builds up in case the inflow rate exceeds the road segment's flow capacity (or service rate). A major difference to the static approach is that travel costs depend on the position in the queue: Travelers at the end of a queue experience higher travel times compared to travelers further ahead.

An important point is that marginal social costs have to be computed *mutatis mutandis*, and not *ceteris paribus* (see e.g., Arnott et al., 1993; Button, 2004). That is, the computation of marginal social cost needs to account for the user reactions of all other travelers. In the case of traffic congestion, the immediate effect of an additional traveler is that total congestion costs increase and consequently *ceteris paribus* marginal congestion costs are positive. In contrast, *mutatis mutandis* marginal costs depend on the existing travel alternatives, e.g. alternative modes, routes or departure times:

1. In case for some users there is a travel alternative associated with no additional costs, existing travelers may switch to that alternative. Consequently, traffic congestion is reduced to the original level. Since total experienced travel costs are the same, *mutatis mutandis* marginal congestion costs are zero.

2. In case there is no travel alternative or all travel alternatives are associated with much higher costs, travelers are not able to adjust and traffic congestion will not decrease. Consequently, *mutatis mutandis* and *ceteris paribus* marginal congestion cost are at the same level.

3. In case for some users there is a travel alternative associated with some additional costs, existing travelers may switch to that alternative. Consequently, traffic congestion somewhat decreases. However, the overall effect of an additional traveler is that total experienced travel costs have increased. Consequently, *mutatis mutandis* marginal costs are greater than zero but below *ceteris paribus* marginal costs.

Vickrey's bottleneck model, which can be seen as the archetypical queue model, is widely used by researchers to investigate the trade off between not arriving at the desired time (schedule delay), and being delayed by traffic congestion. The optimal tolling scheme forces travel demand to disperse over time, traffic congestion is eliminated, and system welfare is maximized. The optimal toll is time-variant; with linear schedule delay functions, toll levels first increase linearly from 0 at the beginning of the rush-hour to a maximum value at the desired arrival time, and then decrease linearly with time to 0 at the end of the rush-hour (Vickrey, 1969). The initial model was extended in several analytical studies, e.g. to account for elastic demand (Arnott et al., 1993), simplistic networks (de Palma et al., 2004), heterogeneous travellers (van den Berg, 2011), hypercongestion (Fosgerau, 2015) or household scheduling preferences (de Palma et al., 2015). Levinson and Rafferty (2004) present a pricing approach where transport users are charged based on the marginal congestion externality and – in contrast to standard economic theory – parts of the toll revenues are immediately used to reimburse the delayed travelers. In Levinson

and Rafferty (2004), marginal congestion cost are computed as the bottleneck's service time per vehicle (the inverse of the flow capacity) multiplied with the number of following vehicles that will be queued (until the queue dissipates). Therefore, the marginal cost function is described as triangular, where the toll jumps to the maximum for the first traveler that passes the bottleneck, and then decreases linearly to zero. In their numerical illustration, where travelers are enabled to adjust their departure times in order to react to the tolling scheme, the morning peak is segmented into 5-minute intervals for which the utility for all travelers is averaged. In different tolling approaches different forms of the time-variant toll are tested. The authors find the "first-pays-most" tolling function to be effective regarding the criteria "reduction of travel delay, schedule delay, user cost, social cost and inequity". In contrast, the highest welfare is obtained for a time-variant toll which is consistent with Vickrey (1969). Verhoef (2003) investigates congestion pricing applying a car-following model, i.e. continuously varying driving speeds over time and space, including hypercongestion. Since there is no closed-form solution, a pricing rule is developed which approximates the welfare maximum. The developed pricing rule builds on the Pigouvian taxation principle transferred to the traffic congestion resulting from a car-following model. A numerical model is used to show that applying a tolling scheme based on Vickrey's bottleneck model may result in a lower welfare compared to the developed pricing rule.

Most congestion pricing studies make use of an analytical modeling approach to investigate illustrative case studies with an often simplistic network and transport demand. Simulation-based approaches allow for real world applications with complex networks and population structures. Nagel et al. (2008) use a large-scale transport simulation to investigate the effects of a pricing scheme which investigates user-specific and time-dependent tolls by setting tolls proportionally to the time spent traveling; however, it does not iteratively optimize tolls. In the context of simulation-based evacuation studies, Lämmel and Flötteröd (2009) develop a routing strategy which approximates user- and road-specific marginal social costs and yields a reduction in total travel time. They argue that each ("causing") agent i in a queue delays every following ("affected") agent j in the same queue by a time increment of $d\tau_j$. Their implementation then includes the following approximations:

1. They assume steady state flow conditions, and therefore can approximate $d\tau_j \approx 1/\bar{q}$, where $\bar{q}$ is the steady state flow.

2. They also approximate the number of affected agents by the time difference between i and the dissolution of the queue.

Kaddoura and Kickhöfer (2014), Kaddoura (2015) and Kaddoura and Nagel (2016) take this one step further by replacing approximation 2 by actively identifying the affected agents, and multiplying each affected agent's additional delay by its individual value of travel time savings, before allocating the resulting marginal cost to the causing agent. All these approaches have problems when the true bottleneck is not at the end of the present link but further downstream, and there are different variants in how that is treated; for a detailed description, see Sec. 2.2.1. Agarwal and Kickhöfer (2015) and Kaddoura and Nagel (2018) combine the same approach with a simulation-based methodology to reduce air pollution and noise exposure costs.

One could assume that a more complex iterative traffic simulation requires a more complex computation of the optimal toll levels. For example, a time-dependent traffic simulation might require the computation of time-variant tolls; for a traffic model with user-specific values of travel time savings (VTTS), the computation of optimal toll levels might need to account for the users' variable VTTS; for more complex assumptions regarding the transport users' choice dimensions, optimal tolling might become more difficult since marginal social costs need to be computed *mutatis mutandis*, i.e. accounting for the transport users' behavioral reactions. A plausible way of dealing with more complex traffic simulations is thus to extend the computation of optimal tolls towards a higher level of detail. One approach is to compute link-, time- and user-specific

congestion tolls based on the dynamic queuing effects at bottlenecks (see e.g. Kaddoura and Kickhöfer, 2014; Kaddoura, 2015).

The present study proposes an alternative approach which neglects the complex interaction of transport users at bottlenecks. This rather simple approach computes time- and link-specific congestion prices and corrects these prices from iteration to iteration applying a simple control-theoretical approach. The present paper investigates the potential and effectiveness of the proposed approach to improve the overall system welfare, and discusses advantages compared to more complex optimization approaches. In this study, Vickrey's bottleneck model is transferred to an iterative agent-based simulation approach which allows for large-scale networks and a sophisticated representation of transport demand. The model is used to develop and investigate different congestion pricing approaches. The newly developed congestion pricing approaches are compared to pre-existing congestion pricing approaches (Kaddoura and Kickhöfer, 2014; Kaddoura, 2015). The focus lies on the following two pricing approaches:

1. A rather simple time interval-based based list pricing approach described in Sec. 2.2.2 in which marginal congestion cost and tolls are the same for all travelers in the same time interval.

2. A more complex approach which accounts for the queuing dynamics, i.e. the position in the queue and resulting delay effects, described in Sec. 2.2.1.

Simulation results are compared to Vickrey's original model. Finally, to make full use of the simulation-based methodology, the pricing approaches are applied to a real-world case study of the Greater Berlin area.

2 Methodology

2.1 Transport simulation framework: MATSim

2.1.1 Overview

In MATSim[1], transport users are simulated as individual agents. For each agent, daily travel plans have to be provided which describe the initial behavior, i.e. when to end an activity and how to travel to the next activity location. The agents are enabled to adjust their initial behavior applying an evolutionary iterative approach. In every iteration, (1) the plans are executed (Traffic flow simulation), (2) evaluated (Evaluation) and (3) new plans are generated (Learning).

1. **Traffic Flow Simulation** All agents simultaneously execute their travel plans and consequently interact in the simulated physical environment. Traffic congestion is simulated based on a queue model where each road segment (link) is modeled as *First In First Out* queue (Gawron, 1998) with a free speed travel time, a flow capacity c_{flow}, and a storage capacity (causing spill-back). For a detailed description of the bottleneck model's economic interpretation, see Sec. 2.1.2.

2. **Evaluation** Each agent evaluates the executed plan based on predefined behavioral parameters and utility functions. A plan's utility is typically composed of two parts: (i) a trip-related disutility (e.g. travel time, toll, distance-based cost) and (ii) the utility gained from performing activities. The latter part is based on the approach by Charypar and

---

[1]Multi-Agent Transport Simulation, see www.matsim.org

Nagel (2005) where the marginal gain is typically positive but decreases with the duration spent at the activity location:

$$V_{p,a} = \beta^{perf} \cdot t_a^{typ} \cdot \ln\left(t_{p,a}^{perf} \Big/ t_{0,a}\right) \; , \tag{1}$$

where $t_{p,a}^{perf}$ is the time person p performs activity a, t_a^{typ} is an activity's *typical duration*, β^{perf} is the marginal utility of performing an activity at its typical duration, and $t_{0,a}$ is defined as

$$t_{0,a} = t_a^{typ} \cdot \exp(-1) \; ; \tag{2}$$

see Horni et al. (2016, Sec. 97.4.2) for a discussion of this setting.

An activity can only be performed during activity-specific opening times. If an agent arrives at an activity location before or after the activity is open, the early or late arrival penalty results from the opportunity cost of time which is approximately equivalent to β^{perf}. Additionally, there may be a late arrival penalty β^{late} which comes on top of the opportunity cost of time if an agent arrives after the *latest arrival time* which can be specified for each activity.

3. **Learning** During choice set generation, some agents generate new travel plans by making a copy of an existing plan and changing parts of the copied plan such as the departure time or the transport route ("innovation"). The other agents choose among their existing plans based on a multinomial logit model.

   At some iteration, innovation is switched off, meaning that from then on the choice sets will be fixed, and agents will only use the logit model to select between those plans that are in their respective choice sets.

Repeating these steps allows the agents to improve and obtain plausible travel plans, and the simulation results stabilize. Assuming each agent's travel plans to represent a valid choice set, the outcome is an approximation of the stochastic user equilibrium (Raney and Nagel, 2006; Nagel and Flötteröd, 2012; Horni et al., 2016).

2.1.2 The queue model's economics

To get started, let us look at the economics of a single isolated queue. For a single link and assuming all agents to enter the link simultaneously, the total travel cost function is

$$\begin{aligned}
C(x) &= VTTS \cdot \left(t_{free} \cdot x + \sum_{i=1}^{x} (i-1) \cdot c_{outflow}^{-1} \right) \\
&= VTTS \cdot \left(t_{free} \cdot x + \left(\frac{x^2 - x}{2} \right) \cdot c_{outflow}^{-1} \right),
\end{aligned} \tag{3}$$

where $C(x)$ is the total cost; $VTTS$ is the value of travel time savings[2]; x is the number of agents entering the link; t_{free} is the free speed travel time; and $c_{outflow}$ is the (out-)flow, which here is equal to the (out-)flow capacity since there is no spillback from downstream links. That is, first all x travellers incur the free speed travel time, t_{free}. Then, each departing vehicle consumes a time headway of $c_{outflow}^{-1}$, meaning that the ith vehicle is confronted with an additional waiting time of $(i-1) \cdot c_{outflow}^{-1}$. Evidently, $C(x)$ increases quadratically with the number of car users. Marginal and average cost are

$$MC(x) = \frac{\partial C(x)}{\partial x} = VTTS \cdot \left(t_{free} + \left(x - \frac{1}{2} \right) \cdot c_{outflow}^{-1} \right) \; , \tag{4}$$

---

[2]In this simple illustration the VTTS is assumed to be equal for all agents on the link. However, MATSim explicitly allows for agent-specific behavioral parameters including agent-specific VTTS.

$$AC(x) = \frac{C(x)}{x} = VTTS \cdot \left(t_{free} + \left(\frac{x}{2} - \frac{1}{2} \right) \cdot c_{outflow}^{-1} \right) , \qquad (5)$$

where $MC(x)$ is the marginal travel cost and $AC(x)$ is the average travel cost.

With constant inflow, Eq. (3) becomes

$$C(x) = VTTS \cdot \left(t_{free} \cdot x + \left(\frac{x^2 - x}{2} \right) \cdot \left(c_{outflow}^{-1} - c_{inflow}^{-1} \right) \right) ,$$

where c_{inflow} is the inflow rate or time headway between each agent, and traffic congestion only appears in case $c_{inflow} > c_{outflow}$ (active bottleneck, see e.g. Daganzo, 1997). This just replaces $c_{outflow}^{-1}$ by $(c_{outflow}^{-1} - c_{inflow}^{-1})$ in Eqs. (3) to (5), that is, the functional form of the dependence on x will not be changed.

From these equations, the following can be observed for the queue model:

- At $x = 0$, marginal and average cost are the same. For $x > 0$, marginal cost is larger than average cost which is consistent with the basic (= static) economic theory (see e.g. Maibach et al., 2008).

- In heavily congested situations, the outflow $c_{outflow}$ might become quite small, for example because of spillback from downstream and/or a situation close to a grid lock. $c_{outflow}^{-1}$ thus would become very large, and in consequence the slopes of $MC(x)$ and $AC(x)$ would become close to vertical, and the situation would become similar to a purely capacity-limited facility, as for example described by Button (1993, Chapter 6.2).

2.2 Simulation-based congestion pricing

This section describes the different congestion pricing implementations that will be investigated in this paper. They can be grouped into two types: (i) a queue-based internalization of delays and (ii) (time) interval based list prices.

2.2.1 Queue-based congestion pricing (QCP)

The QCP methodology computes delays at an agent-specific level based on the queuing dynamics at the bottleneck link. That is, an agent's toll depends on the position in the queue and the number of following travelers (in the same queue) as well as their VTTS.

In each iteration, marginal congestion costs are estimated based on the *ceteris paribus* effect, i.e. neglecting the agents' reactions of using alternative routes or times of the day. However, from iteration to iteration the agents are enabled to change their travel behavior in order to avoid traffic congestion or toll payments. That is, over the entire simulation process, i.e. for multiple iterations, user reactions are accounted for, and the computation of marginal congestion cost moves towards the *mutatis mutandis* principle (see e.g., Arnott et al., 1993; Button, 2004).

Pre-existing QCP approach   The pre-existing congestion pricing methodology allocates delays to the causing agent. The basic mechanism is described below. For a detailed description of this approach, see Kaddoura and Kickhöfer (2014), Kaddoura (2015) and Kaddoura and Nagel (2016).

- For each agent leaving the link r ("affected agent"), the link-specific delay is computed as the difference between free speed travel time and actual travel time.

6

- If an agent's delay is > 0 sec, iterate through link r's queue (starting with the first agent ahead and then proceeding with the next agent ahead, and so on) and let each "causing" agent pay for $1/c_r^{flow}$ **until the delay is fully internalized (cost recovery)** or until the vehicle initially starting link r's queue is reached (meaning there is spill-back resulting from a downstream link; see next point).

- Spill-back related delays are taken into account as follows: For each causing agent ahead who pays for $1/c_r^{flow}$, the same amount of time is deducted from the affected agent's total delay which needs to be internalized. Without any spill-back related delays, there is no remaining delay after iterating through link r's queue. In case the delay is not fully internalized after iterating through link r's queue, the non-internalized delay is stored with the affected agent. The agent will carry it into the next link, and attempt to allocate it there to the vehicles ahead, if there is a queue at the end of that next link. This may eventually lead to the spill-back causing link, where the bottleneck is active, i.e. the link's flow capacity prevents agents from moving to the next link.

The pre-existing methodology can be considered as marginal congestion pricing with price-cap, resulting in toll revenues that correspond to the level of the congestion cost (cost recovery). Consequently, the pre-existing approach may underestimate optimal toll levels.

New QCP approach (a)   This approach is similar to the pre-existing one but goes beyond cost recovery. Instead, each agent in the queue ahead is considered as the causing agent. In contrast to the pre-existing approach, this approach may result in toll revenues that exceed total congestion costs.

- If the delay is > 0 sec, iterate through link r's queue and let **each** agent ahead pay for $1/c_r^{flow}$.

- Spill-back related delays are taken into account as follows: Non-internalized delays after iterating through the link r's queue are added to the agent's delay on the next downstream link, as in the pre-existing QCP approach.

New QCP approach (b)   This approach is similar to QCP approach (a), except that spill-back related delays are **not** additionally taken into account.

Some discussion of QCP approaches   One notes that the description of QCP approaches becomes quite involved. There are also many design decisions that need to be made, for example:

- Should causing agents be charged $1/c_r^{flow}$, or their actual time headway? The latter feels more intuitive, but would be grossly unfair if the actually active bottleneck would be further downstream (and the causing agent maybe even take a different route). – The present investigation always allocateds $1/c_r^{flow}$.

- Should affected agents receive compensation for their delay, or should the toll go to some authority? The former might lead to some "bucket brigade" (Holland, 1992) flow of compensations where, say, each agent could charge the agent ahead *fully* for its delay, and the agent ahead could then charge the agent further ahead for its own delay, thus only having to pay the difference. – The present investigation takes the second alternative.

- Should delays be allocated upstream across nodes? Clearly, the actually active bottleneck might be further downstream, pointing to a "yes" answer. However, for this the algorithm would have to identify the actually active bottleneck. – The present investigation makes certain compromises here, as explained above.

2.2.2 Interval-based list pricing (LP)

The LP approach computes tolls based on the congestion level in a previous iteration. These tolls, or interval-based list prices (see e.g., Schlechte and Tanner, 2010; Tanner and Mitusch, 2011), are computed for each link and time interval. In contrast to the QCP methodology, all agents traveling on the same link and within the same time interval face the same toll payments. As control loop feedback mechanism two different controllers, a rather simple control approach (Controller A) and a proportional-integral-derivative (PID) controller (Controller B) are embedded in the iterative simulation approach described in Sec. 2.1. Both controllers use road charges to control traffic and to decongest the transport network.

1. Initialization: Run the simulation for n^{pre} iterations without any price setting.

2. Compute the average delay per link and time interval as

$$d_{r,t,k}^0 = \frac{\sum_n^{N_{r,t,k}} \left(t_{r,t,n,k}^{act} - t_r^{free} \right)}{N_{r,t,k}} \ , \tag{6}$$

   where r denotes the link, t is the time interval, k is a counter for the price adjustment iteration, n denotes the agent, $N_{r,t}$ is the total number of agents leaving link r in time interval t, $t_{r,t,n,k}^{act}$ is the actual travel time and t_r^{free} is the free speed travel time.

3. A link and time interval is considered as congested if a certain threshold value is exceeded. Thus, set the average delay which is processed by the controller to

$$d_{r,t,k} = \begin{cases} 0 & \text{for } d_{r,t,k}^0 < d^{min} \\ d_{r,t,k}^0 & \text{for } d_{r,t,k}^0 \geq d^{min} \ , \end{cases} \tag{7}$$

   where d^{min} denotes the threshold value.

4. Control:

   LP Controller A: Controller A adjusts the price per link and time interval as follows:

$$m_{r,t,k} = \begin{cases} m_{r,t,k-1} + a & \text{for } d_{r,t,k} > 0 \\ max\{0, m_{r,t,k-1} - a\} & \text{for } d_{r,t,k} = 0 \ , \end{cases} \tag{8}$$

   where $m_{r,t,k}$ denotes the adjusted toll per link and time interval, $m_{r,t,k-1}$ is the previous toll, and a is the toll adjustment value.

   LP Controller B: Controller B is a discrete PID controller, the price per link and time interval is adjusted as follows:

$$m_{r,t,k} = max\{0, K_p \cdot d_{r,t,k} + K_i \cdot D_{r,t,k} + K_d \cdot (d_{r,t,k} - d_{r,t,k-1})\} \ , \tag{9}$$

   where K_p, K_i and K_d denote tuning parameters, $D_{r,t,k}$ is the current integral value, $d_{r,t,k-1}$ is the previous average delay. $D_{r,t,k}$ is initially set to zero and then changed as follows: If $d_{r,t,k}$ is positive, $D_{r,t,k}$ is increased by $d_{r,t,k}$. If $d_{r,t,k}$ is zero, $D_{r,t,k}$ is decreased by deducting the unused road capacity which is computed as $w \cdot \left(\frac{T}{N_{r,t,k}} - \frac{1}{c_r^{flow}} \right)$, where T is the time bin size of time interval t and w is a scaling factor.

5. Keep the tolls constant and let the transport demand adjust for k^{update} iterations. Thereby, enable all agents to adjust their choice sets for $k^{update} \cdot k^s$ iterations, where k^{update} is the toll update interval and k^s is the fraction of iterations during which agents are enabled to explore new travel alternatives. For the final $(1 - k^s) \cdot k^{update}$ iterations, the agents are only allowed to select from their existing choice sets based on a multinomial logit model.

6. Repeat step 2 and 3 several times.

7. Finally, run the simulation for n^{post} iterations without adjusting the prices.

3 Illustrative example: Vickrey's bottleneck scenario

3.1 Scenario setup

This section describes how Vickrey's bottleneck scenario (Vickrey, 1969; Arnott et al., 1990, 1993) is applied in the agent-based and dynamic context of MATSim.

The **network** consists of three subsequent links link 1, link 2 and link 3, where link 2 is the bottleneck with a flow capacity of 1800 vehicles per hour (see Fig. 1). Spill-back effects are neglected. The free speed travel time from link 1 to link 3 amounts to 202 sec.

$$(0) \longrightarrow (1) \longrightarrow (2) \longrightarrow (3)$$
$$\text{link 1} \qquad \text{link 2} \qquad \text{link 3}$$

Figure 1: Illustrative scenario: Network

The **demand** is modeled as 7200 agents that each perform one trip from link 1 to link 3. The behavioral parameters α, β and γ which in Arnott et al. (1990) denote the marginal cost of traveling, early arrival and late arrival are set as follows: $\alpha = 12$, $\beta = 6$, $\gamma = 18$ in monetary units per hour. The desired arrival time is 8.30 a.m. by setting both the activity opening and latest activity start time on link 3 to the same value. Each agent's utility is computed based on the travel time cost, the schedule delay cost and the monetary tolls. Further utility components are neglected. Demand elasticity is introduced by allowing for departure time choice. The total number of iterations is set to 500. In each iteration 5% of all agents are enabled to randomly shift their departure times within the range of 2 hours. All other agents select from their existing choice sets according to a multinomial logit model. The choice set is limited to 5 plans per agents. Further parameters are set as follows: $k^{update} = 1$ and $k^s = 1.0$, meaning tolls are updated in every iteration; $T = 300$ sec; $d^{min} = 30$ sec; $a = 1.0$ monetary units; $K_p = K_i = K_d = 0.01$; and $w = 10.0$ (for an explanation of these parameters, see Sec. 2.2.2).

3.2 Results

Tab. 1 provides the total travel time, delay, user costs (including toll payments), toll revenues and system welfare for the final iteration in the base case (no pricing) and the different congestion pricing approaches. All congestion pricing approaches are found to reduce total travel delay and

Table 1: Vickrey's bottleneck scenario.

	Base case (no pricing)	LP controller A $a = 1.0$	LP controller B $K_p=K_i=K_d=0.01$	QCP pre-existing approach	QCP approach (a) & (b)
Total travel time [hours]	4,289	459	440	3,035	767
Total delay [hours]	3,885	55	36	2,631	363
Difference in user benefits including toll payments [monetary units]	n/a	−24,221	−18,725	−16,141	+19,199
Toll revenues [monetary units]	0	65,335	57,902	31,570	20,829
Difference in system welfare [monetary units]	n/a	+41,114	+39,177	+15,429	+40,028

to increase system welfare. The LP approach (both controllers) and the QCP approaches (a) and (b) result in a higher increase in system welfare compared to the pre-existing QCP approach; note that QCP (a) and (b) are identical for this setup since there is no spillback.

Fig. 2 depicts the number of departing and en route agents per time of day (in 5-minute time bins) for the final iteration in the base case and the different congestion pricing approaches. This

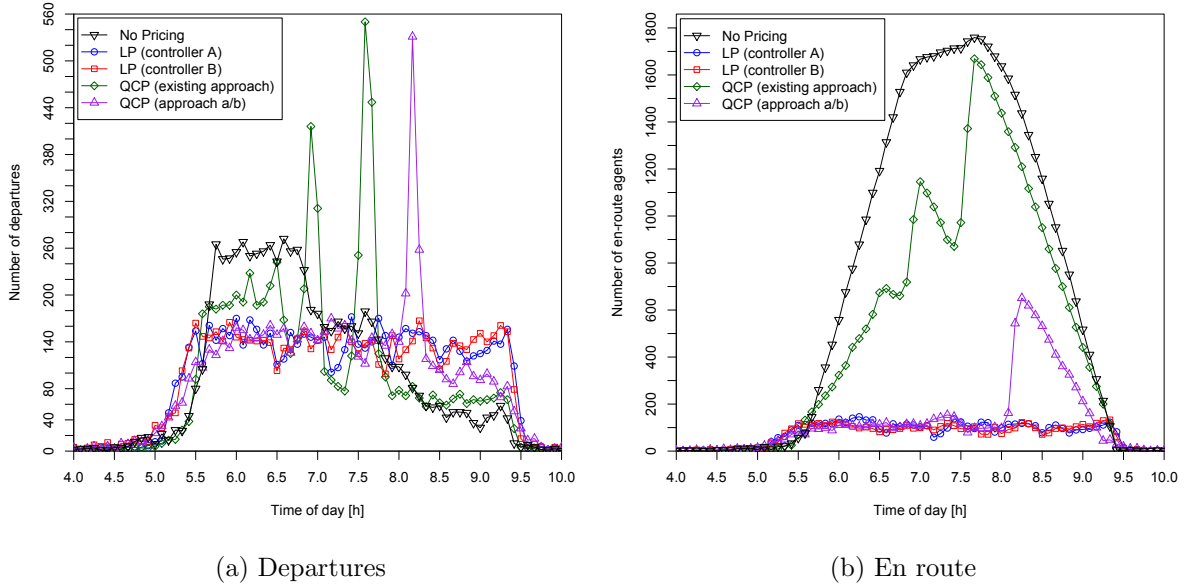(a) Departures   (b) En route

Figure 2: Number of departing and en route agents per time of day

visualization reveals that in all congestion pricing experiments the number of en route agents is at an overall lower level compared to the base case. The LP approach (Controller A and B) approximates a uniform distribution of departures and en route agents during the morning period. In contrast, for the QCP approach the number of en route agents and departure time distribution depicts one (QCP approaches (a) and (b)) or several (QCP pre-existing approach) peaks.

Fig. 3 depicts the toll level per time of day for the final iteration in the different congestion pricing experiments. In the LP approach, the toll time distribution is triangular shaped. The toll first
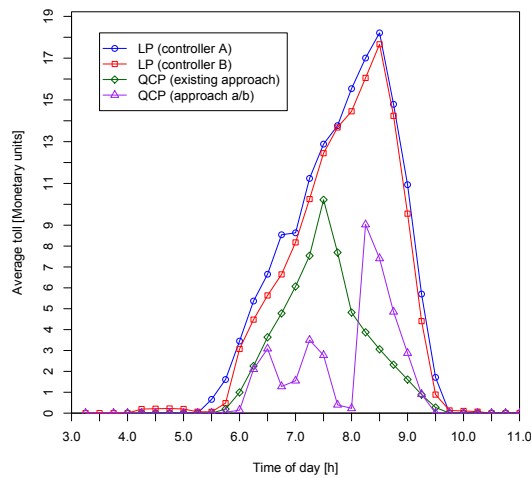

Figure 3: Average toll per time of day.

increases from 0 at around 5.30 a.m. to a maximum value of approximately 18 monetary units at the desired arrival time (8.30 a.m.) and then decreases to 0 at around 9.45 in the morning. In

the pre-existing QCP approach, the average toll per time of day is at an overall lower level; with a maximum value of 10 monetary units at around 7.30 in the morning. The QCP approaches (a) and (b) result in an overall lower toll level compared to the LP approach, with several peaks.
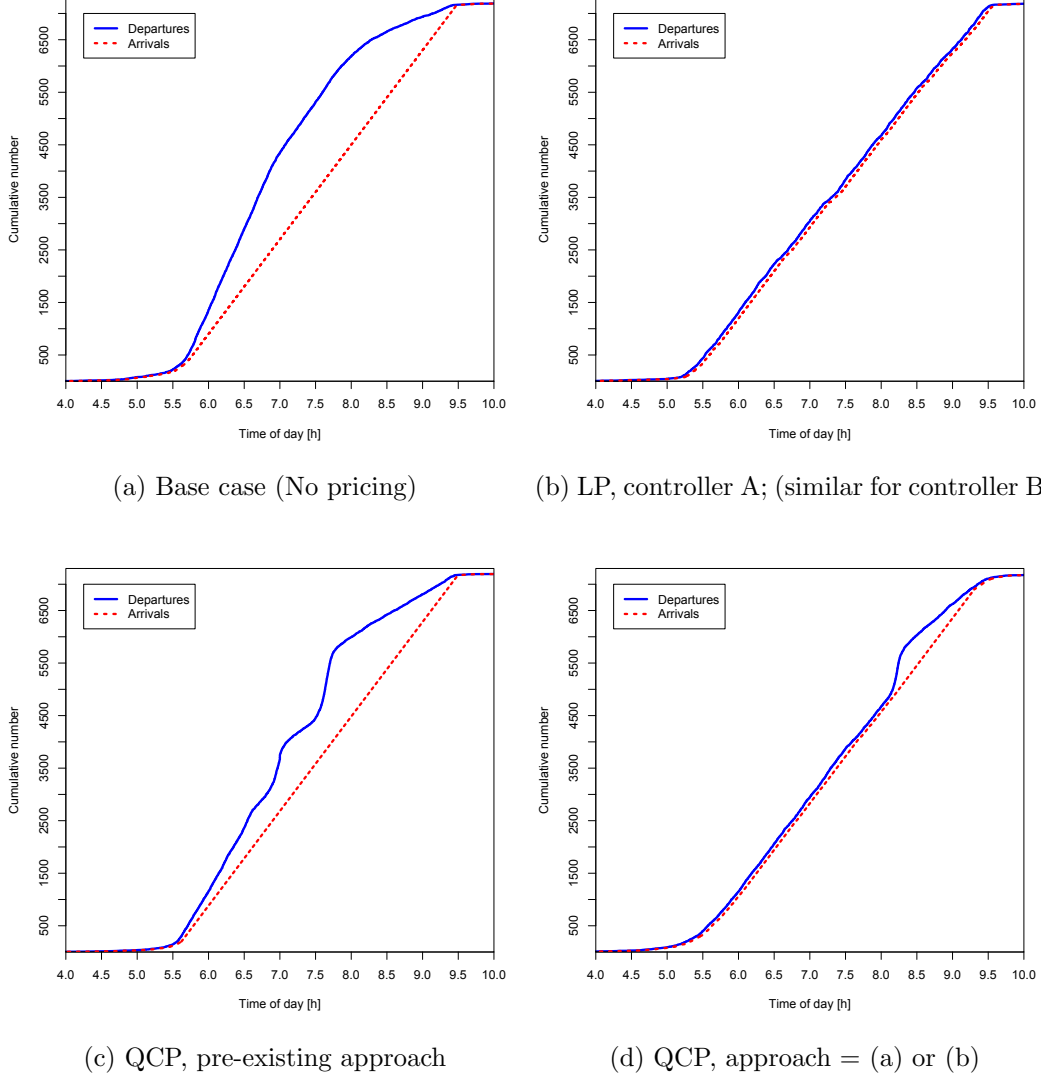

(a) Base case (No pricing)   (b) LP, controller A; (similar for controller B)

(c) QCP, pre-existing approach   (d) QCP, approach = (a) or (b)

Figure 4: Cumulative departures and arrivals in the final iteration

In Fig. 4, the cumulative departures and arrivals are shown for the morning period. The horizontal distance between cumulative departures and arrivals describes the travel time, and the vertical distance describes the queue length. Thus, the area between the two curves indicates the level of traffic congestion which is observed to be very high in the no pricing situation, reduced for the QCP approach and very low applying LP approach.

The following lessons can be taken from these results:

- The pre-existing QCP approach is not strong enough to bring congestion down to zero (see Fig. 4c). This corresponds to the fact that the pre-existing QCP approach only charges up to cost recovery, whereas economic theory would demand to charge marginal cost, which is larger.

- The QCP (a) or (b) approaches have stability issues in the Vickrey scenario. The reason is that with QCP the first vehicle in a queue always pays most. In consequence, the early

positions in the queue become very unattractive. The drivers react by underutilizing the capacity during these early times. The system reacts by repeatedly "breaking" the queue, thereby reducing the charges for the early vehicles (see the "QCP a/b" lines both for the toll in Fig. 3 and for unused capacity in Fig. 5). Note how, in Fig. 3, QCP a/b at least
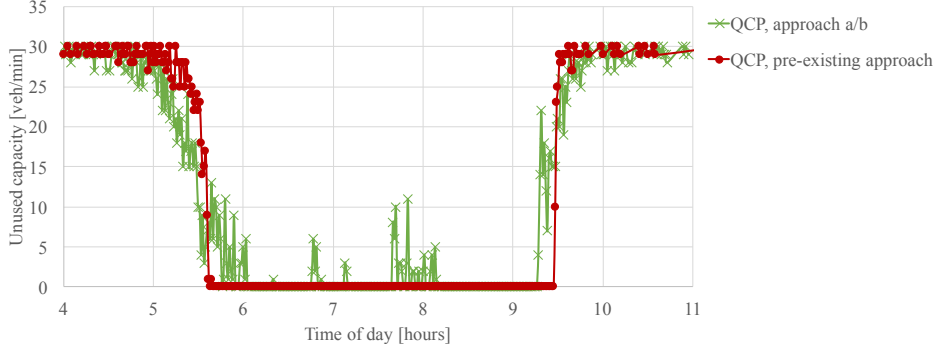

Figure 5: The bottleneck's unused capacity per time of day.

gets the position of the maximum approximately right. However, it clearly has problems getting the ramping up to that level right.

4 Real-world case study: Berlin, Germany

4.1 Scenario setup

The congestion pricing approaches are applied to a real-world case study of the Greater Berlin area generated by Ziemke et al. (2018) based on the methodology developed by Ziemke et al. (2015). The transport **supply** consists of all major and minor roads in the Greater Berlin region, including all public transit lines. In Ziemke et al. (2018), the transport **demand** is modeled as individual agents which amount to 10% of the population in the Greater Berlin region. To model realistic congestion effects, the road segments' flow and storage capacities are accordingly reduced. Simulated traffic volumes are calibrated against hourly traffic counts for a total of 346 traffic count stations in the Berlin area. Furthermore, the simulation results are validated using SrV 2008 survey data (Ahrens, 2009) accounting for, e.g., the total trip number, departure time distribution, trip distances and travel times. Each agent has an individual activity schedule and is differently pressed for time. Consequently, the resulting VTTS is different for each transport user (Kaddoura and Nagel, 2016; Charypar and Nagel, 2005). In the base case, the average VTTS per person amounts to 7.57 EUR/h with a 5% percentile of 4.53 EUR/h, a median of 6.74 EUR/h and a 95 % percentile of 16.31 EUR/h. In this study, the total number of iterations is set to 500. Only during the first 400 iterations choice sets are generated and the agents are enabled to generate a new travel plan based on the enabled choice dimensions (see Sec. 4.2).

4.2 Simulation experiments

For different assumptions regarding transport users' choice dimensions, a selection of pricing schemes is applied to the Greater Berlin case study. In a first setup, transport users are only enabled to change their routes. In a second setup, transport users are enabled to adjust their routes, departure times and modes of transportation.

Each choice dimension is represented by a strategy which is either enabled or disabled.

- **Route choice:** A new transport route is generated based on the least cost path taking

into account mean travel times and toll payments from the previous iteration(s). That is, the agents may accept a longer route, i.e. higher distance- or time-based travel costs, in order to avoid toll payments.

- **Departure time choice:** Activity end times are randomly shifted within a maximum range of 2 hours. That is, the agents may accept schedule delay costs and deviate from the desired activity schedule in order to avoid traffic congestion or toll payments.

- **Mode choice:** For each sub-tour (i.e. trips starting and ending at the same activity location), the transport mode is randomly set to only car, only bicycle or a random combination of public transit or walking.

Note that, in MATSim, the strategies generate new alternatives. The actual selection of the alternatives is performed in a separate choice model, based on the scores of the executed plans.

During the first 400 iterations, each strategy is applied by 5% of the transport users. All other agents select from their existing choice sets according to a multinomial logit model. In the final 100 iterations, choice set generation is disabled and all agents select from their existing choice sets. Each agent's choice set is limited to 5 plans which is considered to yield a plausible level of fluctuations in daily travel behavior from one iteration to the next one. The LP parameters are set as follows: $k^{update} = 1$ and $k^s = 1.0$, meaning tolls are updated in every iteration; $T = 900$ sec; $d_{min} = 1$ sec. In contrast to Vickrey's illustrative bottleneck example, the real-world Berlin case study requires a computation time of several days. Hence, to avoid a large number of parametric runs, the number of tuning parameters is reduced. K_i and K_d are set to 0, and instead the method of successive averages is used to smoothen the toll levels over the iterations.

4.3 Results

Resulting toll payments: Each pricing approach results in different toll payments. Tab. 2 provides the average and maximum toll payments per car trip for each pricing and simulation setup. The total toll which is paid during a trip results from the congestion level on each road

Table 2: Resulting toll payments

	LP controller B $K_p = 0.003$ $K_i = K_d = 0$	QCP pre-existing approach	QCP approach (a)	QCP approach (b)
Average toll per car trip [EUR]				
Route choice only	0.31	0.19	0.55	0.46
Route, mode and departure time choice	0.27	0.16	0.39	0.34
Maximum toll per car trip [EUR]				
Route choice only	3.88	4.11	12.62	12.36
Route, mode and departure time choice	3.63	4.78	12.25	9.27

segment along the transport user's route from one activity to the next one. For a single road segment, toll payments may be different depending on the time of day (QCP, LP) or position in the queue (QCP). Tab. 2 points out that even though there is no upper limit, toll levels stay within a certain range because transport users are not willing to accept large payments and rather take a detour or switch to an alternative time or mode. This, in turn, reduces the congestion level on that road segment in that time bin which in turn reduces the toll level. In the LP approach, for example, the maximum toll per road segment (= per bottleneck) amounts to 1.75 EUR in the simulation experiment with route choice only and 1.25 EUR in the simulation experiment

with route, mode and departure time choice. A comparison of the different simulation setups reveals that with mode and departure time choice, average toll payments are at a lower level compared to the route choice only case.

Aggregated analysis: Tab. 3 and Tab. 4 provide the changes in travel time, travel distance, delay, user benefits (including the negative effect of toll payments), toll revenues and system welfare of each congestion pricing approach compared to the base case (no pricing).

Table 3: Berlin: Route choice only (upscaled to 100%); comparison with the base case (no pricing)

	LP controller B $K_p = 0.003$ $K_i = K_d = 0$	QCP pre-existing approach	QCP approach (a)	QCP approach (b)
Change in car travel time [thousand hours]	-49.40	-32.57	-41.61	-43.55
Change in average travel time per car trip [sec]	-23	-15	-19	-20
Change in car travel distance [thousand km]	512.32	210.97	983.38	753.21
Change in average travel distance per car trip [km]	0.06	0.03	0.12	0.10
Change in car delay [thousand hours]	-64.85	-39.45	-75.43	-68.43
Change in average delay per car trip [sec]	-30	-18	-34	-31
Change in user benefits including toll payments [thousand EUR]	-2,192.26	-1,263.34	-4,251.34	-3,441.66
Toll revenues [thousand EUR]	2,451.33	1,466.62	4,359.18	3,611.05
Change in welfare [thousand EUR]	259.07	203.28	107.84	169.39

All pricing experiments are found to reduce total car travel time and delay (direct congestion effects). The LP approach yields the largest increase in system welfare and reduction in average travel time.

The pre-existing QCP approach yields lower prices compared to the LP approach. Consequently, the remaining level of congestion is larger compared to the LP approach. In contrast, the QCP approaches (a) and (b) yield a higher level of congestion prices compared to the LP approach. The reduced overall welfare gains imply that tolls here are now so high that the network utilization is below the optimum.

In the route choice only case, the total travel distance slightly increases which indicates that transport users take longer detours in order to avoid congested and/or tolled road segments. In contrast, with mode and departure time choice, the average travel distance per car trip increases even though the total car travel distance decreases. This indicates that short-distance travelers switch from car to alternative modes, while long-distance travelers remain within or switch to the car mode. Similar, the reduction in average travel time per car trip is lower compared to the simulation experiments without mode and departure time choice.

The comparison of the different choice dimensions reveals that mode and departure time choice increase the overall positive effect of the pricing schemes. With mode and departure time choice the increase in system welfare which results from the LP approach multiplies by two compared to the only route choice case.

Table 4: Berlin: Route, departure time and mode choice (upscaled to 100%); comparison with the base case (no pricing)

	LP controller B $K_p = 0.003$ $K_i = K_d = 0$	QCP pre-existing approach	QCP approach (a)	QCP approach (b)
Change in number of car trips (in thousands)	-554.04	-341.07	-986.31	-797.43
Change in car travel time [thousand hours]	-293.93	-187.14	-460.91	-388.03
Change in average travel time per car trip [sec]	-14	-11	7	-3
Change in car travel distance [thousand km]	-3,492.73	-2,115.76	-6,488.22	-5,182.00
Change in average travel distance per car trip [km]	0.54	0.33	0.98	0.78
Change in car delay [thousand hours]	-118.36	-81.82	-146.84	-134.75
Change in average delay per car trip [sec]	-48	-33	-58	-54
Change in user benefits including toll payments [thousand EUR]	-1,450.14	-834.71	-2,409.67	-1,996.38
Toll revenues [thousand EUR]	1,968.39	1,245.55	2,705.48	2,379.25
Change in welfare [thousand EUR]	518.26	410.84	295.81	382.87

Spatial analysis – Traffic congestion: The spatial analysis confirms the above described reduction in traffic congestion. Fig. 6 depicts an inner-city motorway bottleneck situation and the resulting queues in the base case and the pricing policy (LP; Controller B; simulation experiment with route, mode and departure time choice).

Spatial analysis – Traffic volumes: Fig. 7 and Fig. 8 depict the changes in traffic volumes resulting from the pricing policy (LP, Controller B) in the simulation setup without and with mode and departure time choice. The simulation setup without mode and departure time choice reveals an overall tendency of reduced demand levels on major roads including the inner-city motorway and increased demand levels on minor roads (see Fig. 7). In contrast, the simulation setup with route, mode and departure time choice, traffic volumes are at an overall lower level (see Fig. 8) since transport users switch from car to alternative modes.

Person- and trip based mode shift analysis: In Fig. 9, the changes in transport modes are shown for each person and trip in the base case and the policy case (LP, Controller B; simulation setup with route, mode and departure time choice). Fig. 9 depicts all trips including those of transport users remaining within their previous transport mode. Fig. 9 reveals that for most trips, transport users remain within their previous transport mode and only a relatively small share of trips switch from car to bicycle (5%) and public transit (3%).

(a) Map section

(b) Bottleneck situation

(c) Base case (no pricing); 8.45 a.m.

(d) Pricing (LP, Controller B); 8.45 a.m.

Figure 6: An inner-city motorway bottleneck situation: Red indicates a delayed vehicle; Green indicates a non-delayed vehicle. Map layer: © OpenStreetMap contributors. (Simulation experiment with route, departure time and mode choice)
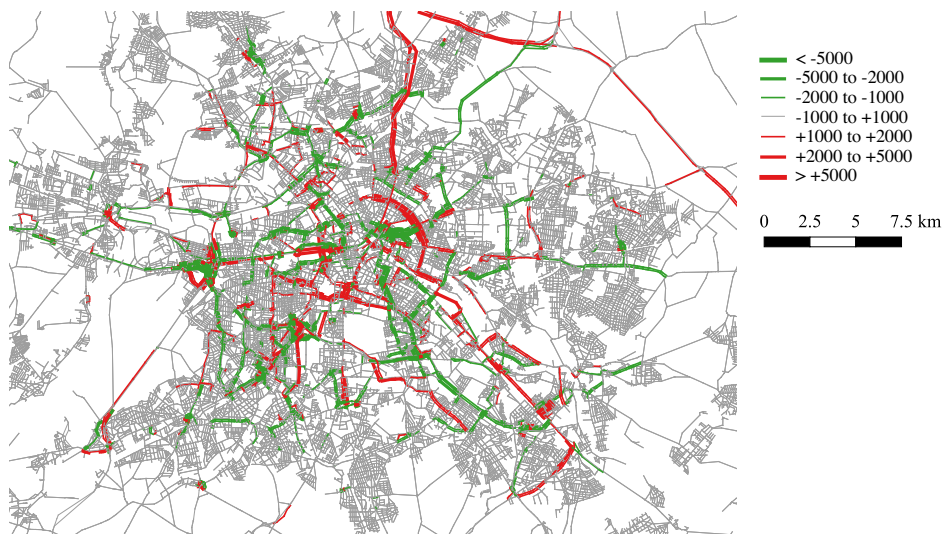

Figure 7: Changes in daily traffic volume resulting from the pricing scheme (LP, Controller B) only allowing for route choice. Red road segments indicate an increase in traffic; Green road segments indicate a decrease in traffic.
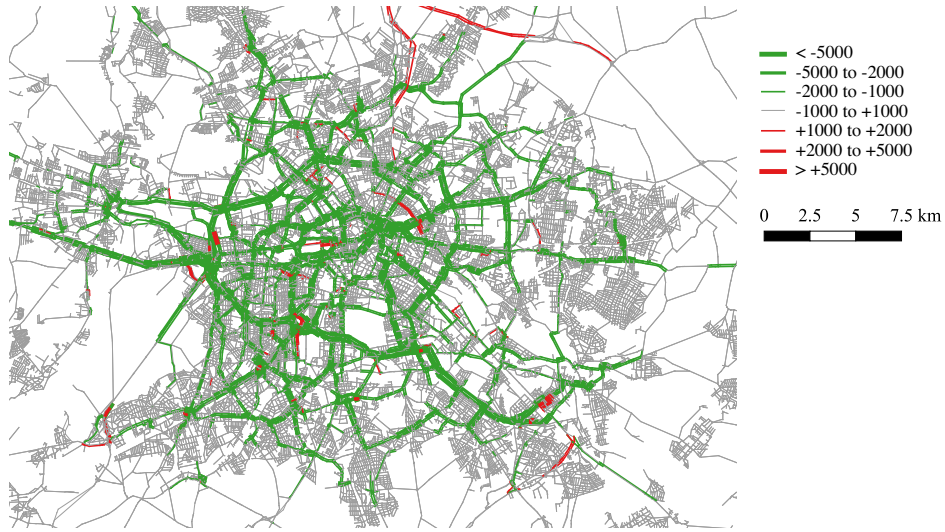
Figure 8: Changes in daily traffic volume resulting from the pricing scheme (LP, Controller B) allowing for route, departure time and mode choice. Red road segments indicate an increase in traffic; Green road segments indicate a decrease in traffic.
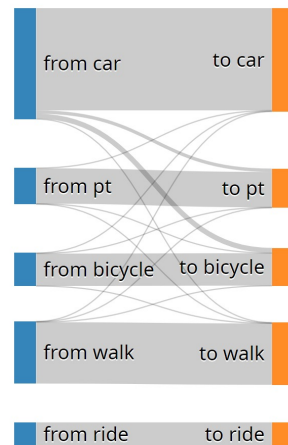

Figure 9: Person- and trip-specific mode switch analysis: Congestion pricing (LP, Controller B) vs. base case; Simulation experiments with route, mode and departure time choice

5 Discussion

The QCP approach internalizes delays directly, with each person in the queue paying for the additional delay it is causing to each person later in the queue. This has the immediate consequence that the first person in the queue has to pay most. On the other hand, it is known from, e.g., Vickrey (1969) or Arnott et al. (1990), that the optimal toll in this situation is triangular, i.e. with the first person in the queue paying little, the toll then increasing to a maximum, and then decreasing again. The reason, as Arnott et al. (1990) explain, is that internalization has to happen not *ceterus paribus*, as the QCP approach does, but *mutatis mutandis*, i.e. taking into account behavioral reactions. This results in unstable behavior: At some point, early travellers avoid tolled routes because they are too expensive. However, at this point very early travellers can use it again, since the queue will be broken at some point and thus the toll will be affordable. This entices more and more travellers to use the tolled route, until at some point the queue is no longer broken, the early travellers pay a lot, and the adaptive cycle is started again.

One might thus have speculated that QCP is approximately optimal when departure time choice and mode choice are switched off, since then the *mutatis mutandis* reaction does not seem possible. However, the simulation experiments reveal that also without mode and departure time choice, LP results in a higher system welfare compared to the QCP approach. An explanation for this may be that even without mode and departure time choice, the circular behavior described in the previous paragraph can occur, and thus the system remains strongly fluctuating, making it difficult for the adaptive agents to find a good solution. Conversely, for the real world scenario in Sec. 4 with departure and mode choice enabled, the QCP approach is not as bad as one might have expected based on these thoughts. Presumably, any kind of internalization leads to improvements. Still, the LP approach is not only the most successful one here, but also to prefer because it is the simplest of the approaches that are investigated.

Both applications of the developed pricing rules show that delays are significantly reduced and the system welfare is increased. However, there is a certain amount of delay which remains. For LP-B and $K_i = K_d = 0$, there indeed has to be some remaining delay, since with these parameters Eq. (9) implies that delay $d = m/K_p$, where m is the toll and K_p is a positive constant. Increasing the penalty for traveling in a congested time period, e.g. by setting a (LP; Controller A) or K_p, K_i, K_d (LP; Controller B) to very high values further reduce or even eliminate traffic congestion. This, however, reduces the system welfare which is explained by a too strong reaction of the agents to avoid the congestion charges. With such drastic tolls, over many iterations, the agents maximize their individual utility by rather traveling far too early or far too late (and deviating from individually optimal activity durations) in order to avoid paying high tolls. Thus, departing during or even close to the peak time becomes unattractive.

Controller B allows for an economic interpretation. The economic theory suggests that optimal prices should reflect marginal external costs (see e.g. Maibach et al., 2008). A proportional controller ($K_i = K_d = 0$) where travel delay is considered as the error value and where $K_p = VTTS$ may be interpreted as average congestion cost pricing. Taking into account the travel cost functions described in Sec. 2.1.2, K_p in the range of approximately twice the average $VTTS$ may approximate a marginal congestion pricing approach. And indeed, our average VTTS, see Sec. 4.1, is 7.57 $EUR/h \approx 0.0021\ EUR/sec$, meaning that twice that value is similar to the value of 0.003 that was found heuristically. More theory, however, would be needed to fully untangle the relation between queueing, behavioral reaction, and toll levels.

The LP approach offers two additional advantages over the QCP approach. First, the computation of congestion prices is much easier, which results in a better computational performance. Second, the LP approach computes link-specific tolls for a predefined time interval length, e.g. 1 hour, that are equal for all transport users within the same time interval. This makes the real-world application of the LP tolling scheme more feasible compared to the QCP approach

where each transport user pays a different amount depending on her or his position in the queue.

Nevertheless, the LP tolling scheme only provides a first step towards the design of welfare improving pricing policies that are applicable to the real-world.

The mechanism to model transport users' departure time adaptation which is applied in the Berlin and Vickrey case study randomly shifts activity end times within a certain range. The adjusted activity schedule is then tested and evaluated in the next iteration(s) (see Sec. 2.1). Since the maximum number of travel plans per agent is limited, the worst plan will be discarded. Each travel plan is then re-evaluated in the following iterations, i.e. every time it is chosen from the existing choice set and executed. That is, the learning mechanism accounts for the travel plans' reliability (see e.g., Neumann et al., 2016). Nevertheless, the applied mechanism to model transport users' departure time adaptation follows a rather simple approach. For a more complex time adaptation approach, which applies several search rules and explicitly accounts for, e.g., uncertainty, search costs and different sources for the knowledge about departure time-specific travel conditions, see, e.g., Xiong and Zhang (2013).

6   Conclusion

This paper investigates optimal congestion pricing strategies using an agent-based simulation framework which allows for real-world application and complex user behavior. Traffic congestion is simulated based on a queue model, where limited space on roads may cause spill-back and move bottlenecks upstream. Transport users iteratively adjust their travel behavior based on their experienced travel cost. Hence, the applied simulation approach goes beyond the simplifying assumptions typically made in other models in order to reduce the real-world complexity. Instead, the applied simulation approach accounts for iterative learning, stochastic user behavior, and only approximates the user equilibrium, which may be considered as closer to real-world than a model where transport users behave completely rational, have a perfect knowledge about all travel alternatives, and travel behavior strictly follows the user equilibrium.

This study investigates the no closed-form solution of the optimization problem of how to set link-, time- and user-specific road charges in order to increase the overall system welfare under the above described circumstances. The literature provides some help to address this problem: Following the concept of Pigouvian taxation, road charges are optimized by setting them equivalent to the (link-, time- and user-specific) marginal external costs. By doing so, experienced travel costs are corrected and individual travel decisions are changed towards an overall higher level of efficiency. The resulting changes in the transport system depend on the assumptions regarding how transport users can react to the road charges. In the case of departure time choice and queue-based bottleneck dynamics, the literature suggests to expect traffic congestion to completely dissolve (see e.g. Vickrey, 1969).

Unfortunately, defining the correct marginal external costs is not obvious in a time-dependent model with a queue-based traffic flow behavior. In this paper, two pricing rules are developed and investigated. The first one directly builds on the Pigouvian taxation principle and computes marginal external congestion costs based on the queuing dynamics at the bottleneck links; resulting toll payments differ from agent to agent depending on the position in the queue (QCP approach). The second one builds on the desired outcome and uses control-theoretical elements to adjust toll levels depending on the congestion level in order to reduce or eliminate traffic congestion; resulting toll payments are the same for all agents in the same time bin and on the same road segment (LP approach).

The pricing rules are applied to Vickrey's bottleneck model and the case study of the Greater Berlin area. The simulation experiments reveal that with and without departure time and mode choice, the rather simple LP rule results in a higher system welfare compared to the more

complex, and seemingly more "correct", QCP approach. In the end, the QCP approach only approximates optimal toll levels since it computes external congestion costs *ceteris paribus*, i.e. based on the travel behavior in the current iteration, and not *mutatis mutandis*, i.e. taking into consideration the transport users' reactions. The rather complex calculation of optimal toll levels, i.e. accounting for the queuing dynamics, appears to result in large fluctuations from iteration to iteration which leads to a less stable simulation outcome. In contrast, the LP rule appears to better take into account the system's dynamics and the agents' learning behavior and produces a more stable outcome. The results also reveal that the pricing rules significantly reduce traffic congestion, however, there is still a remaining delay, even in the simulation experiments with mode and departure time choice.

Overall, this paper points out further need for research and contributes to the exploration of optimization heuristics for real-world oriented simulation approaches which allow for a complex user behavior and a high level of realism. This paper also contributes to the practical implementation of pricing policies. The simulation results indicate the importance to account for the system's dynamics and implement a robust pricing scheme which smoothens the day-to-day traffic fluctuations, e.g. by rather slowly adjusting time-specific toll levels from one time period (day/week/month) to the next one.

Acknowledgments

References

Agarwal, A. and B. Kickhöfer. Agent-based simultaneous optimization of congestion and air pollution: A real-world case study. *Procedia Computer Science*, 52(C):914–919, 2015. ISSN 1877-0509. doi: 10.1016/j.procs.2015.05.165.

Ahrens, G.-A. Endbericht zur Verkehrserhebung Mobilität in Städten – SrV 2008 in Berlin. Technical report, Technische Universität Dresden, 2009. http://www.stadtentwicklung.berlin.de/verkehr/politik_planung/zahlen_fakten/download/2_SrV_endbericht_tudresden_2008_berlin.pdf.

Arnott, R., A. de Palma, and R. Lindsey. Economics of a bottleneck. *Journal of Urban Economics*, 27 (1):111–130, 1990. ISSN 0094-1190. doi: 10.1016/0094-1190(90)90028-L.

Arnott, R., A. de Palma, and R. Lindsey. A structural model of peak-period congestion: A traffic bottleneck with elastic demand. *The American Economic Review*, 83(1):161–179, 1993. ISSN 00028282. URL http://www.jstor.org/stable/2117502.

Button, K. *Transport economics*. Edward Elgar Publishing Limited, 2nd edition, 1993.

Button, K. The Rationale for Road Pricing: Standard Theory and Latest Advances. *Research in Transportation Economics*, 9(Supplement C):3–25, 2004. ISSN 0739-8859. doi: 10.1016/S0739-8859(04)09001-8.

Charypar, D. and K. Nagel. Generating complete all-day activity plans with genetic algorithms. *Transportation*, 32(4):369–397, 2005. ISSN 0049-4488. doi: 10.1007/s11116-004-8287-y.

Daganzo, C. *Fundamentals of Transportation and Traffic Operations*. Pergamon, 1997. ISBN 9780080427850. URL https://books.google.de/books?id=FihPAAAAMAAJ.

Fosgerau, M. Congestion in the bathtub. *Economics of Transportation*, 4(4):241–255, 2015. ISSN 2212-0122. doi: 10.1016/j.ecotra.2015.08.001.

Gawron, C. An Iterative Algorithm to Determine the Dynamic User Equilibrium in a Traffic Simulation Model. *International Journal of Modern Physics C*, 9(3):393–407, 1998.

Holland, J. *Adaptation in Natural and Artificial Systems*. Bradford Books, 1992. Reprint edition.

Horni, A., K. Nagel, and K. W. Axhausen, editors. *The Multi-Agent Transport Simulation MATSim*. Ubiquity, London, 2016. doi: 10.5334/baw.

Kaddoura, I. Marginal Congestion Cost Pricing in a Multi-Agent Simulation: Investigation of the Greater Berlin Area. *Journal of Transport Economics and Policy*, 49(4):560–578, 2015.

Kaddoura, I. and B. Kickhöfer. Optimal Road Pricing: Towards an Agent-based Marginal Social Cost Approach. VSP Working Paper 14-01, TU Berlin, Transport Systems Planning and Transport Telematics, 2014. URL http://www.vsp.tu-berlin.de/publications.

Kaddoura, I. and K. Nagel. Agent-based congestion pricing and transport routing with heterogeneous values of travel time savings. *Procedia Computer Science*, 83:908–913, 2016. doi: 10.1016/j.procs.2016.04.184.

Kaddoura, I. and K. Nagel. Simultaneous internalization of traffic congestion and noise exposure costs. *Transportation*, 45(5):1579–1600, 2018. doi: 10.1007/s11116-017-9776-0.

Lämmel, G. and G. Flötteröd. Towards system optimum: Finding optimal routing strategies in time-tependent networks for large-scale evacuation problems. In Mertsching, B., M. Hund, and Z. Aziz, editors, *KI 2009: Advances in Artificial Intelligence*, volume 5803 of *LNCS (LNAI)*, pages 532–539. Springer, Berlin, 2009. doi: 10.1007/978-3-642-04617-9_67.

Levinson, D. M. and P. Rafferty. Delayer Pays Principle: Examining Congestion Pricing With Compensation. *International Journal of Transport Economics*, 31(3):295–311, 2004.

Maibach, M., D. Schreyer, D. Sutter, H. van Essen, B. Boon, R. Smokers, A. Schroten, C. Doll, B. Pawlowska, and M. Bak. Handbook on estimation of external costs in the transport sector. Technical report, CE Delft, 2008. URL `http://ec.europa.eu/transport/sites/transport/files/themes/sustainable/doc/2008_costs_handbook.pdf`. Internalisation Measures and Policies for All external Cost of Transport (IMPACT).

Nagel, K. and G. Flötteröd. Agent-based traffic assignment: Going from trips to behavioural travelers. In Pendyala, R. and C. Bhat, editors, *Travel Behaviour Research in an Evolving World – Selected papers from the 12th international conference on travel behaviour research*, pages 261–294. International Association for Travel Behaviour Research, 2012. ISBN 978-1-105-47378-4.

Nagel, K., D. Grether, U. Beuck, Y. Chen, M. Rieser, and K. Axhausen. Multi-agent transport simulations and economic evaluation. *Journal of Economics and Statistics (Jahrbücher für Nationalökonomie und Statistik)*, 228(2+3):173–194, 2008. URL `http://www.digizeitschriften.de/dms/resolveppn/?PID=PPN345616359_0228%7CLOG_0020`. URL `http://www.digizeitschriften.de/dms/resolveppn/?PID=PPN345616359_0228%7CLOG_0020`.

Neumann, A., I. Kaddoura, and K. Nagel. Mind the gap – Passenger arrival patterns in multi-agent simulations. *International Journal of Transportation*, 4(1):27–40, 2016. doi: 10.14257/ijt.2016.4.1.02.

de Palma, A., R. Lindsey, and E. Quinet. Time-Varying Road Pricing and Choice of Toll Locations. *Research in Transportation Economics*, 9(Supplement C):107–131, 2004. ISSN 0739-8859. doi: 10.1016/S0739-8859(04)09005-5.

de Palma, A., R. Lindsey, and N. Picard. Trip-timing decisions and congestion with household scheduling preferences. *Economics of Transportation*, 4(1):118–131, 2015. ISSN 2212-0122. doi: 10.1016/j.ecotra.2015.02.001.

Pigou, A. *The Economics of Welfare*. MacMillan, New York, 1920.

Raney, B. and K. Nagel. An improved framework for large-scale multi-agent simulations of travel behaviour. In Rietveld, P., B. Jourquin, and K. Westin, editors, *Towards better performing European Transportation Systems*, pages 305–347. Routledge, London, 2006.

Schlechte, T. and A. Tanner. Railway capacity auctions with dual prices. Technical Report 10-10, Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB), 2010.

Tanner, A. and K. Mitusch. Trassenvermarktung: Auktion versus Listenpreisverfahren. *Internationales Verkehrswesen*, 63(3):15–19, 2011.

van den Berg, V. A. *Congestion Pricing With Heterogeneous Travellers*. PhD thesis, Tinbergen Institute, 2011.

Verhoef, E. Inside the queue: hypercongestion and road pricing in a continuous time–continuous place model of traffic congestion. *Journal of Urban Economics*, 54:531–565, 2003.

Vickrey, W. Congestion Theory and Transport Investment. *The American Economic Review*, 59(2):251–260, 1969.

Xiong, C. and L. Zhang. A positive model of departure time choice under road pricing and uncertainty. *Transportat Research Record*, (2345):117–125, 2013.

Ziemke, D., K. Nagel, and C. Bhat. Integrating CEMDAP and MATSim to increase the transferability of transport demand models. *Transportation Research Record*, 2493:117–125, 2015. doi: 10.3141/2493-13.

Ziemke, D., I. Kaddoura, and K. Nagel. The MATSim Open Berlin Scenario: An openly available agent-based transport simulation scenario based on Open Data. VSP Working Paper 18-xx, TU Berlin, Transport Systems Planning and Transport Telematics, 2018. URL `http://www.vsp.tu-berlin.de/publications`.