A Dynamic Cordon Pricing Scheme combining a Macroscopic and an Agent-based traffic Models

Nan Zheng
Urban Transport Systems Laboratory
School of Architecture, Civil and Environmental Engineering
École Polytechnique Fédérale de Lausanne
EPFL ENAC INTER LUTS, GC C2 406, 1015-Lausanne, Switzerland
Phone/Fax: +41(21)693-2484 / +41(21)693-2479
E-mail: nan.zheng@epfl.ch

Rashid A. Waraich
Institute for Transport Planning and Systems
Swiss Federal Institute of Technology, Zurich
HIL F 31.3, CH-8093, Zurich, Switzerland
Phone/Fax: +44(46)633-3943 / +44(46)633-1057
E-mail: rashid.waraich@ivt.baug.ethz.ch

Kay W. Axhausen
Institute for Transport Planning and Systems
Swiss Federal Institute of Technology, Zurich
HIL F 31.3, CH-8093, Zurich, Switzerland
Phone/Fax: +44(46)633-3943 / +44(46)633-1057
E-mail: axhausen@ivt.baug.ethz.ch

Nikolas Geroliminis*
Urban Transport Systems Laboratory
School of Architecture, Civil and Environmental Engineering
École Polytechnique Fédérale de Lausanne
EPFL ENAC INTER LUTS, GC C2 389, 1015-Lausanne, Switzerland
Phone/Fax: +41(21)693-2481 / +41(21)693-2479
E-mail: nikolas.geroliminis@epfl.ch

*corresponding author

For Presentation Only at the 91th Annual Meeting
Transportation Research Board
Washington, D.C., January 2012

Words count: 5400 words + 7 Figures + 1Table = 7400 words
ABSTRACT

Pricing is considered an effective management policy to reduce traffic congestion in transportation networks. This paper combines a macroscopic modeling of traffic congestion in urban networks with an agent-based simulator to study congestion pricing schemes. The macroscopic model, which has been tested with real data in previous studies, represents an accurate and robust approach to model the dynamics of congestion. The agent-based simulator can represent the complexity of travel behavior in terms of departure time choice and heterogeneous users. While traditional traffic simulators (including car-following, lane-changing and route choice models) consider traffic demand as input, i.e. inelastic to level of congestion conditions. On the other hand, most of traditional congestion pricing models utilize a network supply curve which is not consistent with the physics of traffic and the dynamics of congestion, as they are sensitive to demand fluctuations and non-stationary conditions. Also, many of the existing pricing models are assuming deterministic and homogeneous population characteristics. In this paper, we first demonstrate by case studies in Zurich urban road network, that the output of a multi-agent based simulator, is consistent with the physics of traffic flow dynamics, as expressed by a macroscopic fundamental diagram (MFD). We then apply a dynamic cordon-based congestion pricing scheme, in which tolls are controlled by an MFD, and investigate the effectiveness of the proposed pricing scheme. Results show that by applying such a congestion pricing, (i) the savings of travel time at both aggregated and disaggregated level outweighs the costs, (ii) the congestion inside cordon is eased while no congestion is shifted to outside cordon, (iii) during toll period, fewer work-related activities shift starting time than leisure-related activities do; while the impact of toll is more significant in the evening than morning. Future work can apply the same methodology to other network-based pricing schemes. Equity issues can be investigated more carefully, if provided with data such as income of agents. Time-dependent or value-of-time-dependent pricing schemes then can also be determined.
INTRODUCTION

To alleviate traffic congestion in cities, congestion pricing has been proposed by researchers and policy makers, with direct applications to cities (e.g. London, Singapore, Stockholm, etc). The intention is to change travellers’ behaviour, such as departure time or route choice and to reduce congestion by charging users for the external costs they create. The vast literature of congestion pricing can be composed in two categories, marginal-cost pricing models and bottleneck models.

A comprehensive literature summary of these models can be found in Yang and Huang [1]. The theoretical background of marginal-cost pricing has relied on the fundamental concept, first introduced by Pigou [2] and followed by Vickrey [3] and other researchers: if on each link of a network a toll is charged, which equals to the additional congestion cost imposed on other users by an extra traveller, the sum of consumer surplus and total revenue is maximized. In the traffic assignment literature tolls of this type belong to the first-best pricing and have been proposed to drive a user equilibrium pattern [4] toward a system optimum. Despite their idealized theoretical basis, first-best pricing models have been impractical and difficult to implement. Merchand [5] investigated second-best tolls using a general equilibrium model. According to the second-best pricing models, e.g. Arnott et al. [6], tolls are charged in a subset of selected links where the bottlenecks are. In all these models, demand elasticity is realized by introducing cost to traveller’s utility. For example, in Vickrey’s model [7] a traveller experiences a delay cost of waiting in the queue and a penalty, “schedule delay”, which is the difference between the actual time passing the destination and the desired time; accordingly, the traveller may adjust his departure time to avoid high schedule delay. Equilibrium is obtained when no individual has an incentive to alter his departure time. However, the common inadequacies of the models are: (i) utility function is identically given to travellers, while in reality travellers vary in desired arrival time, willingness to pay and etc., (ii) demand elasticity is limited to departure time and route choice where in reality mode choice and travel or not travel are also important factors and (iii) Pigouvian-type tolls assume a network supply curve (desired or input demand vs. average travel cost) is not consistent with the physics of traffic [8].

The second issue of pricing, as already mentioned above, is that charging individual links is difficult to implement. Instead, pricing schemes of aggregated links and networks have been developed and applied in different cities [9]. Recently, Maruyama and Sumalee [10] compared performance of cordon- and area-road pricing schemes on their efficiency and equity. Yang and Huang [11] examined the principle of marginal-cost pricing in a road network, Anderson and Mohring [12] examined congestion on the Twin Cities road network having drivers face marginal rather than average costs to reflect optimal prices using a user equilibrium assignment for a single period. The basic ambiguity in most of these models is that traffic conditions are considered stationary. Furthermore, the traditional network supply curve for congestion pricing modelling, relating input demand to average travel cost, is not consistent with the physics of traffic. This is because [13] for a given average flow, i.e. desired demand over a period of time, the total cost expressed in delay terms (i) is sensitive, during congested conditions, to small variations of flow within the given period and (ii) depends on the initial state of the system and the level of congestion. On the other hand, it has been broadly shown through simulation and field experiments [13] [14] that plots between pertinent variables flow, speed and density on a spatially disaggregated level (one link) are very chaotic and do not follow a well-defined curve. The main reason is that traffic systems are not in steady-state conditions at a link level. Thus, the estimated congestion toll based on idealized versions of these curves may not be optimal and the system may
be either still congested if under-priced or very uncongested if over-priced. According to the same research, a Macroscopic Fundamental Diagram (MFD) model can better capture traffic behaviour on an aggregated level, say an urban network, without the detailed knowledge of conditions in individual links.

Recently, Geroliminis and Levinson [8] combined Vickrey’s theory with a macroscopic traffic model to identify the equilibrium solution for a congested network in the no-toll case. A dynamic model of cordon-based congestion pricing (such as for the morning commute) for networks was also developed in consistency with the physics of traffic. In comparison to the bottleneck model, in the network case the optimal length of the toll period was found to be smaller than the congestion period in the no-toll case and the total delay savings to be higher than the total toll paid. While consistent with the physics of traffic, the above work assumes deterministic and homogeneous population characteristics. This might result in non-optimal estimated tolls. Agent-based models are possible solutions for representing demand elasticity and heterogeneity. This is because heterogeneous travelling agents are used in the models: each agent (i) has an individual utility function, (ii) has individual value of travel time savings and (iii) affects other agents’ decisions [9]. If the output of an agent-based model shows the property of the MFD, we can develop a dynamic network-wide congestion pricing schemes controlled by a macroscopic tool. This approach is more robust since tolls are determined based on traffic flow dynamics, rather than the traditional models based on demand-supply curves and marginal cost, which are sensitive to demand fluctuations and non-stationary conditions.

Based on the discussion above, to test the effectiveness of different congestion pricing schemes such as cordon- or area-based pricing, an idea is: if the output of an agent-based model is able to represent collective traffic behaviour as this expressed by the MFD, it would be interesting to utilize this macroscopic tool to develop a congestion pricing scheme in the agent-based model. Advantages of using this macroscopic tool also lie in that it has lower collection and transaction costs than link-based or area-based tolling and is based on traffic models that are readily observable with existing monitoring technologies [13][14]. Therefore, the objectives of this research are: (i) to investigate the existence of the MFD in an agent-based simulator, (ii) to apply an MFD-controlled cordon-based pricing, (iii) to examine the effectiveness and the issue of equity for the proposed pricing scheme. In our study, the utilized agent-based simulator is MATSim (www.matsim.org), a state-of-the-art multi-agents based simulator which was developed jointly by ETH Zurich and TU Berlin.

The rest of this paper is structured as the following: we describe the basic principles of the MATSim simulator in Section 2 while the main features of the MFD model in Section 3. In Section 4 we explain the algorithm of a MFD controlled dynamic cordon-based pricing and we detail the necessary representation of the physics of traffic we require from MATSim. Results of case studies are given in Section 5 where we show the existence of MFD and the effectiveness of the proposed pricing scheme. In the final section we draw conclusions and point out future work.

**MATSIM MULTI-AGENT TRAFFIC SIMULATOR**

We now introduce some basic concepts of an agent-based traffic simulation model, MATSim, which has been widely applied for transport and land use studies e.g. [15][16] and travel behaviour modelling [17][18]. The simulator integrates activity-based demand generation with dynamic traffic assignment. Activity-based demand generation (ABDG) models generate daily activities in sequence and trips connecting these activities for every “agent” in the network. Demand generation thus is embedded in a concept of daily activity demand from which the need for
transport is derived. Random utility theory is used to generate plans of daily activity. Each agent is assigned with his own utility function therefore behavioural differences among the agents are realized. Besides, in the context of ABDG, the entire activity plan (mode choice, departure time choice and the activity sequence) is the unit of decision to iterate route assignments. These are superior to traditional demand generation model. The simulation structure of the system can be summarized as follows [19]: for every agent, one initial activity plan is given. Input data such as population and land use data, as well as network data are processed to generate this initial demand. Each agent usually has only one plan. The selected activity plans are simulated along the timeline in the model representation of the physical world: (i) of loading the agent on the network link at which the previous activity is located at a given departure time, (ii) of moving the agent along a given route through the network, where it interacts with other agents under way, and (iii) unloading the agent from the network at the link of the destination activity. Different from a variety of models for the simulation of car traffic with discrete entities, including car-following models and lane-changing models, the approximation of traffic in MATSim is fulfilled as the following: each road segment is modelled as a First In First Out waiting queue, with a minimum service time of the length of the road segment divided by the maximum travel speed. The maximum number of vehicles that a queue can discharge equals the road capacity, depending on the number of lanes etc. The capacity is thus a predetermined value, as opposed to models with flow dynamics where the actual maximum outflow is influenced by the number of accumulated vehicles (density) and their interactions. The only concept related to flow dynamics integrated into the queuing simulation is a shockwave between vehicles travelling backwards at constant speed in the case of traffic jam discharge, bringing in some notion of kinematic waves represented in their full scale in macroscopic traffic models. In order to compare activity plans, they are evaluated with a measure of general utility, called score, including utilities for activities and penalties for undesired manners. The related scoring function describes the agent’s preferences. The score of a daily activity plan $U_{plan}$ is given by a utility equation, the detail of which can be found in [19]. Given the score, a replanning strategy is selected by probability such as change of mode choice and another plan is created and then executed; or agents just decides which plan to select from its existing memory for the next execution according to a Logit-type probability. This iteration cycle is stopped after the properties of the system fulfil some stopping criterion. Conceptually, the system has to run until the agents cannot significantly improve the score of the executed plans that is when agent-based stochastic user equilibrium is reached.

THE MACROSCOPIC FUNDAMENTAL DIAGRAM

In this section, we describe the important features and findings of the MFD model. Godfrey [20], Daganzo [21][21], and Geroliminis and Daganzo [13][14] proposed and showed that uniformly congested urban networks approximately exhibit a relation between the numbers of vehicles in the network to the network’s average circulation flow (vehicle kilometres travelled divided by the total length of roads). This happens even though the flow-density plots for individual links (known as Fundamental Diagram) exhibit considerable scatter. This is proved by using a micro-simulation of the San Francisco business district in the United States and a field experiment in downtown Yokohama in Japan. They observed that when the highly scattered plots of speed-density from individual fixed detectors were aggregated for a $10$km$^2$ region, the scatter nearly disappeared and points grouped neatly along a well-defined curve. This shows that (i) the MFD is a property of the network itself (infrastructure and control) and not of the demand, i.e. the MFD should have a well-defined maximum and remain invariant when the demand changes both with the time-of-day
and across days; (ii) the throughput of the network, is maximum for the same value of density of vehicles or average speed, independent of the origin-destination tables; (iii) the average trip length for the study region is about constant with time, i.e. the total outflow (trip finish rates)-density curve has the same shape and (iv) the MFD can be estimated accurately and easily using existing monitoring technologies (e.g. detector data, GPS etc.).

Mathematically speaking, the concept of the MFD is the following. Consider a region of a city, which traffic state is described by properties mentioned in the previous paragraph. Then the state of the system, \( n(t) \), is governed by the mass conservation equation: 

\[
\frac{dn}{dt} = I'(t) - o(n(t)),
\]

where \( n \) is the number of the vehicles in a road network, \( I'(t) \) is the inflow to the network at time \( t \), and \( o \) is the total outflow from the system as a function of \( n \). This equation simply explains that traffic systems are dynamic and to estimate the state of the system at time \( t \), the knowledge of the input flow is not sufficient, but boundary conditions are needed, i.e. the state of the system at a prior time \( t' \). Thus, a traffic model that estimates the average travel time based on a specific demand-cost curve ignores not only variations in the demand, but more importantly that this travel time will be different if the initial state of the system is in different traffic regimes. Theoretically, three traffic regimes can be observed for a unimodal-shape flow-density curve. Regime I represent under-saturated states where queues are transient and the total number of vehicles served is smaller than the maximum possible. Regime II represents saturated states. The links are filled part way with permanent queues and many traffic signals operate at capacity. There is a limit to vehicle accumulation corresponding to queues that fill the links. In this regime, flow is constant, but never larger than the quantity \( \sum L \cdot g \cdot s \), where \( L \) is the link length, \( g \) is the duration of green phase and \( s \) the saturation flow of the signal of an intersection. Furthermore, the critical density is achieved during Regime II. The critical density is the density at which maximum outflow or the capacity is achieved and usually it is the objective for a successful traffic management strategy. Regime III, flow decreasing with accumulation, corresponds to oversaturated states and long queues or spillbacks are observed in many links. These states cannot arise by increasing the input flow, but a restriction from downstream is necessary, for example if queues from downstream links block the departures during the green phase. Regime III consists of states where queues fill the links, vehicles are stopped or moving at saturation flows. Congestion would be unevenly distributed over the network if states of individual link in regimes I and III occurred simultaneously. This would create points beneath the curve, as we will show later. Also note that a reduction of the maximum outflow can be observed. The explanation is: traffic behaves as a queue through a bottleneck and generates spill-backs. Traffic flow departing the queue may not stay at its maximum because vehicles in the queue could not travel fast enough to retain capacity headways. An example of the three regimes is illustrated later.

As pointed out in the previous section, the agent-based simulator MATSim does not focus on modelling in details disaggregated characteristics of traffic, like car-following, queue dynamics etc. This is the main motivation for us to carry out an interesting investigation to see if the outputs of MATSim are consistent with the physics of traffic at an aggregated level, as expressed by a MFD. Apart from investigating the MFD, we also need to examine if spill-back phenomenon can be observed in MATSim since spill-back is the key reason of the existence of capacity decrease. It is known that a model with point queues (where vehicles take zero space), will produce a much slower (or not-existent) decreasing part of MFD (regime III).
METHODOLOGY

Development of an MFD-controlled Cordon Pricing Scheme

Assume that the MFD and other traffic phenomena are well produced in MATSim, we would like to develop a macroscopic pricing scheme controlled by MFD. The idea here, instead of using the traditional demand-supply curve, is to determine a toll so that the targeted network operates at its saturated (maximum throughput) level or at under-saturated level. These two levels correspond to Regime I and II described in previous section. In other words, once traffic state drops to Regime III at time $t$ and the traffic density of the network $K_t$ exceeds the critical density $K_{cr}$, we put a toll to shift agents’ back to Regime II or I. Nevertheless, the centre of a city experiences higher level of congestion than the periphery, so pricing should be applied in a smaller cordon for equity issues. The way of determining when, where and how much a toll should be charged is the following:

(i) Define a cordon area for charge and run the initial simulation until equilibrium is achieved.

(ii) Obtain MFD and network density $K_t$ of the cordon area, and observe when $K_t$ exceeds critical network density $K_{cr}$.

(iii) A certain amount of toll is applied and a new simulation is run until equilibrium is achieved. Now an MFD and a $K_t$ curve after pricing are obtained. We repeat (ii) and then

(iv) Apply a Proportional Controller to update the amount of toll based on equation (1).

\[ Toll_i = Toll_{i-1} + c(K_t - K_{cr}) \]  

The proportional controller is a classic linear feedback control strategy that has been widely-applied in freeway traffic flow management. Detailed information on the controller can be found in [22]. Basically our controller tells us that if there exists $K_t$ exceeds $K_{cr}$, the additional toll being charged is proportional to the difference between $K_t$ and $K_{cr}$. For practical consideration, time index $t$ here refers to morning and evening peak and we estimate two different values of tolls accordingly. Parameter $c$ is the constant proportion, which influences the rate of achieving the optimal toll. $K_{cr}$ is a constant as well and estimated from a MFD directly, recalling that MFD is a property of the network. The algorithm of the pricing scheme is summarized and displayed in Figure 1.

Figure 1 the structure of the proposed congestion pricing schemes controlled by the MFD ("plan" means the activity plan of agents, when tolls are given, agents adjust their plans accordingly)
Data Derivation

To implement the macroscopic traffic analysis, one needs to obtain space mean density (vehicle per kilometer) $K_t$ and trip completion rate (outflow) $A_t$ of a network for certain time intervals. $K_t$ is the number of vehicles in a network, $N_t$, divided by the total length of the network $L$ at time $t$. Since the number of vehicles on one link equals to $k_i \cdot L$, $K_t$ actually is a weighted space mean density and therefore reasonably reflects traffic state at an aggregated level. Trip completion rate (outflow) $A_t$ is simply the sum of the number of finished trips $a_i$ reaching destination at link $i$ for the whole network. See equation (3).

$$K_t = \frac{N_t}{L} = \frac{\sum_i (k_i \cdot l_i \cdot n_i)}{\sum_i (l_i \cdot n_i)} \quad k_i = \frac{e_i - q_i}{l_i \cdot n_i}, \quad A_t = \sum_i a_i, \quad (3)$$

where $i \in I$ denotes individual link $i$ in the network. $q_i$ is the number of vehicles leaving link $i$ while $e_i$ is the number of vehicles entering. $k_i$ is the traffic density of link $i$, $l_i$ is the length of link $i$, $n_i$ is the number of lanes of link $i$.

Network Partitioning

One of the main characteristics for a well-defined MFD is that the target network is homogeneously loaded and congestion is not unevenly distributed. For a big city with congestion heterogeneously distributed, Ji and Geroliminis [23] developed an algorithm for partitioning the network into regions that each with a well-defined MFD. The motivation is to facilitate effective control for each region, as by aggregating congested and uncongested regions, the resulted MFD might exhibit high scatter and is not clear to distinguish the problematic region. We apply the same idea here and utilize a simpler filter to find a centre network where a cordon-based pricing can be easily implemented. To do this, we first calculate density of every individual link of a network. We examine the distribution of these densities and zoom in to the region where congestions exist and the variance of densities during peak periods is below a threshold.

CASE STUDIES AND RESULTS

We carry out our investigation in the simulation of Zurich urban road network and present analysis for three scenarios in this section. Scenario 1 is uncongested, with low demand loaded over a network containing urban arterial roads, distributors and some access roads [24]. While in Scenario 2 and 3, higher demand are employed over a navigational network in which all the links of the studied area Zurich are included [19]. Comparing to Scenario 1, more agents are sent into the network therefore more traffic are generated in Scenario 2 and 3. Comparing Scenarios 2 and 3, lower values of link capacities are applied in order to create additional congestion.

On Traffic Phenomena

The Macroscopic Fundamental Diagram

For a 4.5km-radius centre region of Zurich in Scenario 1, two MFDs are shown in Figure 2a. They show the existence of Regime I and the beginning state of Regime II, indicating this network on a macroscopic level is not heavily congested. This is further confirmed after filtering the network to a 1.5 kilometre area, as shown in Figure 2b. To observe a complete MFD with congested states, we look at Scenarios 2 and 3. In these scenarios, many more agents are employed for simulations and consequently more traffic is generated. The MFDs are shown in Figure 2c and d. We observe that
(i) comparing to Scenario 1, Scenarios 2 and 3 have a complete Regime II in which network is operated at its capacity and (ii) congestion Regime III exists, where the more agents queue in the network while less reach destination. Note that the shapes of the MFD are different due to the differences in the given link capacity in each simulation.

The *Fundamental Diagram (FD) and Spill-back Effect*

We now provide a detailed investigation on how congestion propagates at link level in MATSim. Flow rate \(q_i\) and density \(k_i\) are calculated for individual link \(i\). We study if spillback effects are present and long queues can decrease the output of upstream links. Figure 3 shows the FDs for four consecutive links along one of the arterial roads of Zurich in Scenario 1. The colour of the scatters in the figure corresponds to the link with the same colour. From the figure, we see at density around 10, the flow rate reaches its maximum and remains the same value until density around 30. Then flow decreases as density increases, indicating congestion happen at the latter point. Secondly, we see that congestion spills back from downstream to upstream, as the links in green and yellow operate in Regime I and II while the links upstream experience Regime III. These observations make clear that queues are growing from downstream to upstream and blocking effects are present in the simulator. We also observe that queues propagate from downstream to upstream, as congestion appears upstream at a later time. It is also clear that the individual fundamental diagrams exhibit high scatter, especially in the congested regime.
An Explanation of Scatter in the MFD

Now let us have a more careful look to Scenario 1 for an interesting observation in the shape of its MFD. By zooming in Figure 2a, we find loops in the timely connected scatters. Similar phenomena have been observed by [25] [26] and defined as the hysteresis phenomena: higher network flows are observed for the same average network density in the onset and lower in the offset of congestion. This is because there are different spatial distributions of congestion, high variance among densities, for the same level of average network density for different times of a day. In Figure 4 we draw link density distributions of all links in the network at three different times, where the network holds the same amount of vehicles. The chosen times are 6h35 (the highest point), 9h15 (the lowest point) and 11h40 (medium point) in Figure 2a. They refer to the onset of the morning peak from 6h30 to 9h30, and the offset from 9h30 to 11h30. The size of the “bubble” is the value of density and X-Y axis are coordinates of links. It is clear that traffic is more uniformly distributed at 6h35 than at 11h40, while at 9h15 densities are extremely high at some locations. The standard deviations of density are 24, 29 and 32 veh/km respectively.

Figure 3 an illustration of congestion propagation along consecutive links

Figure 4 density distribution of the network at time 6h35, 9h15 and 11h40. Axis represent coordinates of links (x,y-axis represent coordinates)
On Pricing

A Cordon-based Pricing Controlled by MFD

All the findings so far are consistent with previous studies with simulation data or real life experiment. We now combine the MFD model with the agent-based model MATSim to test the effectiveness of the pricing scheme proposed in Figure 1. The data of Scenario 2 are used, which experiences higher scatter. In this way, we expect to provide a stricter test for the effectivenes/robustness of a cordon-based pricing scheme, given that our macroscopic approach is less accurate. The targeted Cordon area is the closed area inside the red ring, shown in Figure 5. Agents who cross the border of the area, the red line, will pay for a toll. Before we apply any toll on the cordon, we run a simulation in MATSim until it achieves equilibrium. Density over time of the initial run with no pricing is plotted in Figure 6a, while the MFD of the initial run in Figure 6b. Given the “no pricing” curve in these two figures, the following information can be identified approximately and taken as input to equation 2: the critical density is 28veh/km; the periods for charging a toll from 7:30am to 9am and from 4pm to 8pm; an initial toll of 1euro for the morning while 4 euro for the afternoon, and a trial and error value of 1 for parameter c.

We follow the algorithm described in Figure 1. The optimal pricing is achieved after four updates: a 2-euro is charged for the morning peak, while a 10-euro is charged between 18:30 and 19:30 and an 8-euro is charged for the rest part of the evening peak. From traffic point of view, we see in Figure 6a, with increasing of pricing, traffic drops; and in Figure 6b, congestion states in Regime III is shifted to the left part of the MFD. Thus, an aggregated approach for pricing, which does not consider individual link behaviour produces the desired results to reduce congestion and identifies the appropriate value of pricing to meet the goals. Furthermore, we look at the same graphs for the concentric part between the cordon area and outside-cordon area (a 2.5km-radius). The motivation is that the traffic condition for the outside-cordon nearby area should not be significantly decayed. Figure 6c shows that the density of the outside cordon slightly increases, but still in the uncongested regime. The explanation is agents who travelled through the cordon now
choose to detour in order to avoid the toll. But if we look at the MFD for this area, which is shown in Figure 6d, the entire area is operating in Regime I and II.

Table 1 summary of the social of the proposed pricing

<table>
<thead>
<tr>
<th></th>
<th>TTT Savings (In cordon)</th>
<th>TTT Savings (Out cordon)</th>
<th>Effectiveness Ratio</th>
<th>Total Toll</th>
<th>Net Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggregated Social Gain</td>
<td>6356.2 veh-hours</td>
<td>-4540.8 veh-hours</td>
<td>1.5</td>
<td>€23662</td>
<td>€5000/day</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From economic point of view, travel time savings at aggregated level (vehicle hours travelled) for the cordon area should be higher than the extra travel time for the outside cordon area plus the total toll charged. Results are summarized in Table 1. Considering the average value of travel time savings (VTTS) of the agents is 18 Swiss Franc per hour [27], the total gain for our pricing scheme is a positive 5000 euro/day. Note that many congestion charging methodologies charge a toll equal to the delay cost. In our approach, the benefits are even higher as delay savings are higher than total cost paid. Furthermore, we look at the improvement at disaggregated level, which is calculated as vehicle hours travelled over vehicle kilometres travelled. We see that the effectiveness ratio between savings and costs is 7.5, much higher than the ratio between total savings and costs.
savings and total cost which is around 1.5. This indicates that (i) a low amount of additional delay which generates almost no impact for the outside cordon creates significant savings inside the cordon and (ii) traditional cost-benefit analyses procedures can estimate aggregate value that have no physical sense. This idea is supported by many researchers (e.g. [28]) that state that the value of time should be a non-linear function with travel delay, i.e. the effect of a longer delay (in $/min) is much more significant than the effect of a short delay.

**A Deep Investigation of Behavioral Shifts**

In this part, we analyze the impact of pricing on shifting behaviour of agents. Given the available data from MATSim, we classify activities into two groups: purpose of going-to-work and leaving-from-work, as work-related activities (WA); and purpose of going-to-leisure as non-work-related activities (NWA). Figure 7 shows the comparison of time shift of WA and NWA in no pricing and final pricing scenarios. The green bars indicate the toll charge periods. Both groups see a shift in travel behaviour. Agents performing WA tend to shift starting time to earlier while performing NWA tend to shift to both earlier and later. The explanation is for work activities the penalty of earliness is lower than of lateness. 5% WT shift during morning toll period while 15.7% for NWA. For evening toll period, 16.1% WA shift while 19.5% for NWA. We can observe that WA are much less sensitive to small toll, as adjusting WA is much less flexible than NWA and agents are more willing to pay toll to avoid inflexibility. For high toll, the impacts are all significant. The percentiles also show that morning shift is less than evening shift. The explanation is because activity starting time is fixed in the morning, therefore agents have to be on time to avoid penalty of doing nothing; while in the evening it is more flexible. Also, in final pricing there are fewer NWA outside toll period which is due to mode change. A deeper investigation on behavioural shift can be done, if data such as income of agents are available, on the impacts of our pricing scheme on agents with different value of time savings, as inspired by the work of Axhausen et al. [27].

![Figure 7 time shift of WA (a) and NWA (b) in no pricing and the final pricing scenarios](image)

**CONCLUSIONS AND FUTURE WORK**

In this work, we introduce a new method to develop cordon-based congestion pricing scheme. The idea is to utilize a macroscopic traffic model, the Macroscopic Fundamental Diagram, to determine optimal tolls for an urban network. The methodology was evaluated with an agent-based model for the city of Zurich. We first investigate the feasibility of combining an agent
based model with an MFD network representation by examining the outputs of three MATSim simulation scenarios on urban road network of Zurich. We find the productions of the agent-based model are consistent with the physics of traffic. On microscopic level, it is able to reflect spill-back effect once congestion is formed. While on macroscopic level, MATSim shows an MFD between network trip completion rate and network density. In addition, the cause of some scatter in the MFD can be explained by the uneven distribution of congestion, results which are consistent with empirical observations. Given these interesting findings, we test the proposed simple cordon-based pricing in MATSim, utilizing a Proportional Controller to update tolls. By an iterative process (in total of four steps), the congested part Regime III of the MFD for the 1km-radius cordon area is eliminated. Meanwhile, there is almost negligible impact on the outside cordon area (between 1km- and 2.5 km-radiuses). The total travel time savings by imposing the toll outweighs the total paid toll plus the increased travel time for the outside cordon area. The savings of time per distance travelled outweighs significantly the extra cost per distance travelled. Furthermore we examine the behavioural shift on performing work and non-work activities. Non work-related activities tend to change activity time more to avoid tolls, while work trips also show time shift but are more stable to high tolls.

We would conclude that the proposed pricing scheme is effective both from traffic and economic point of view. Determining tolls controlled by MFD is a promising methodology for congestion pricing. Following this work, future studies are recommended on: (i) investigate how other macroscopic pricing schemes can improve mobility patterns, such as area-based pricing or distance based pricing, (ii) if provided data, e.g. income, car ownership, address behavioural shift issues with equity, (iii) investigate the effectiveness of a time-dependent toll and a VOT-dependent toll. The fast convergence of the agent-based model in identifying the optimal toll (4 updates) shows that this methodology could be directly applied by cities without the need of time-consuming and expensive surveys in a large part of the population. Nevertheless, this info can be useful to create more equitable schemes.

REFERENCES


