

Network traffic flow optimization under performance constraints

Alfréd Csikós^a, Themistoklis Charalambous^b, Hamed Farhadi^{c,d}, Balázs Kulcsár^c, Henk Wymeersch^c

^a*Institute for Computer Science and Control, Hungarian Academy of Sciences, Budapest, Hungary. E-mail: csikos.alfred@sztaki.mta.hu.*

^b*Department of Electrical Engineering and Automation, School of Electrical Engineering, Aalto University, Espoo, Finland.*

^c*Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden.*

^d*John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA.*

Abstract

In this paper, a model-based perimeter control policy for large-scale urban vehicular networks is proposed. Assuming a homogeneously loaded vehicle network and the existence of a well-posed Network Fundamental Diagram (NFD), we describe a protected network throughout its aggregated dynamics including nonlinear exit flow characteristics. Within this framework of constrained optimal boundary flow gating, two main performance metrics are considered: (a) first, connected to the NFD, the concept of average network travel time and delay as a performance metric is defined; (b) second, at boundaries, we take into account additional external network queue dynamics governed by uncontrolled inflow demands. External queue capacities in terms of finite-link lengths are used as the second performance metric. Hence, the corresponding performance requirement is an upper bound of external queues. While external queues represent vehicles waiting to enter the protected network, internal queue describes the protected network's aggregated behaviour.

By controlling the number of vehicles joining the internal queue from the external ones, herewith a network traffic flow maximization solution subject to the internal and external dynamics and their performance constraints is developed. The originally non-convex optimization problem is transformed to a numerically efficiently convex one by relaxing the performance constraints into time-dependent state boundaries. The control solution can be

interpreted as a mechanism which transforms the unknown arrival process governing the number of vehicles entering the network to a regulated process, such that prescribed performance requirements on travel time in the network and upper bound on the external queue are satisfied. Comparative numerical simulation studies on a microscopic traffic simulator are carried out to show the benefits of the proposed method.

Keywords: Traffic control; traffic flow; perimeter control; network fundamental diagram; travel time; Quality of Service.

1. Introduction

Urban traffic congestion has become a major issue, since it results in - among others - delays, pollutant emissions, higher energy expenditure and accidents (see, e.g., Bigazzi and Figliozzi (2012) and references therein). Intelligent transportation systems via *control* and *estimation* of traffic flows has been of vital importance to support urban traffic management in order to appropriately use finite road capacity under different traffic conditions.

One efficient urban traffic coordination approach is to adapt traffic lights at signalized intersections. To address this problem, several methods has been deployed at different hierarchical levels (ranging from intersection to network level), e.g., Papageorgiou et al. (2003). Among these methods, advanced urban traffic control is one of the most important techniques aiming at describing urban vehicular networks by some traffic models and then based on these mathematical abstractions to develop (optimal) control solutions. Towards this end, the concept of Network Fundamental Diagram, NFD, often called Macroscopic Fundamental Diagram (MFD), has been adopted as a basis for the derivation of traffic control strategies (e.g., Leclercq et al. (2014)). The theory was first proposed in Godfrey (1969) and further developed in Daganzo and Geroliminis (2008) and Helbing (2009) (its application to experimental data is analyzed in Mahmassani et al. (1987); Geroliminis and Daganzo (2008); Ampountolas and Kouvelas (2015)). Geroliminis and Sun (2011) further investigated what the properties that a network should satisfy are, so that an MFD with low scatter exists, by again using data from a field experiment in Yokohama (Japan). It is concluded that if two traffic states from two different time intervals have the same spatial distribution of link density, then the two time intervals have the same average flow. As a result, the assumption that congestion is evenly distributed across the network made by Geroliminis and Daganzo

(2008) is relaxed. Daganzo (2007) first used the NFD to synthesize a controller that maximizes the network outflow, thus comprising a starting point for using the NFD theory for controlling traffic flow. Several works followed the developed control strategies based on NFD to maximize the capacity of homogeneous traffic networks. In this case, a single-region model with one NFD represents the dynamics of the network appropriately. The paper by Hajiahmadi et al. (2015) formulates the optimal control problem as a mixed integer linear optimization problem, with two types of controllers: perimeter controllers and a switching controller of fix-time signal plans. However, the solution to the problem cannot be used in real time. For alleviating this problem, a Proportional-Integral (PI) controller is proposed by Keyvan-Ekbatani et al. (2012) for real-time gating, with an application to the network of Chania, Greece. By modeling the dynamics of the external queues, a perimeter problem is solved via a Nonlinear MPC formulation in Csikós et al. (2015). Recently, in Haddad and Mirkin (2016), time delays in MFD related control problems have been addressed by means of adaptive control.

Alternative approaches have been used to forecast changing conditions in transportation systems. Due to the complexity of such systems, however, short term travel time estimation and prediction have been in the spotlight for a few decades; see, for example, Vlahogianni et al. (2014) and references therein. We hereby categorise the available techniques according to (1) the regression technique applied (*methodology*), (2) the type of data (*urban or rural*), and (3) the *source of data collection*. First, in order to describe estimation and prediction techniques for travel time, we may follow the model regression methodologies applied. In this vein, both parametric or non-parametric regression techniques have been already suggested. Second, part of the research has been inclined towards the case of the freeway, e.g., Li and Rose (2011) and some to urban Rahmani et al. (2015); Zhan et al. (2013); Jenelius and Koutsopoulos (2013) travel time estimation, or short time prediction. Third, different data sources have been utilized for travel time estimation and predictions (ranging from fixed or mobile sensors to data fusion), see references in Vlahogianni et al. (2014).

The primary goal of the above mentioned works was to inform travellers and hence influence route planning. This means an *indirect* inclusion of the estimated, predicted travel time. *Direct* co-design of travel time estimation/prediction information with urban traffic control solution gives rise to improved transportation service, e.g., Lin et al. (2012) proposes a link

travel time minimization in a predictive way. Ensuring performance (e.g., enforced travel time) metrics via traffic control policies is a relevant, non trivial research path, especially in case of large scale traffic networks. In Yildirimoglu et al. (2015), route choice models under dynamical constraint have been included to perimeter flow control decisions. Note that, as reported in Mahmassani et al. (2013); Yildirimoglu and Geroliminis (2014); Kouvelas et al. (2017), route choice effects can influence the trip length distribution of vehicles in the network and thereby approximation of outflow might experience some errors. In Haddad (2017b) optimal control for maximum queue length inclusion on aggregated inter-regional boundary queues is considered. Furthermore, Haddad (2017a) also deals with perimeter control problems for single-region cities, but it focuses on a dynamic model that decomposes the accumulation into two vehicle conservation equations. Optimal control solution to the problem takes decoupled state constraints into consideration.

In this paper, our main contribution is to develop admission control solutions under multiple performance requirements: i.e., provide an upper bound average network travel time and keep the external queue size below a certain threshold. In contrast to recent works Haddad (2017b,a), in our paper these performance requirements are jointly considered and used by rolling horizon capacity maximization where performance requirement are realtime relaxed into changing upper and lower hard constraints for the internal queue dynamics. While perimeter flow control has recently received a lot of attention from a control theoretic perspective, further performance requirements for the system, such as average travel delay in the network, have not been considered. In this work, similarly to the classical perimeter control problem, the objective is to optimize network performance through the maximization of network throughput. However, we additionally include performance requirements, adopting the service indicators of communication networks (see, for example, Klessig and Fettweis (2014); Liu et al. (2014.); Le et al. (2012) and references therein) to (a) keep the travel time spent in the network below a certain threshold, and (b) avoid, if possible, the blockage at the entrance of external queues.

The above performance requirements are incorporated as constraints into the gating design procedure. The problem emanating from our objective and constraints, is first generally formulated as a constrained optimization problem. Furthermore, the general non-convex optimization problem can be transformed to a convex problem for a single step receding horizon

control. The performance of our approach is demonstrated via a case study and compared to that of the proportional-integrator (PI) controller and the unrelaxed MPC problem. Perimeter control is implemented on a network implemented on Vissim, a microscopic traffic simulator, describing part of the inner city of Stockholm, Sweden.

The rest of the paper is organized as follows. In Section 2 we provide the necessary notation and preliminaries for the development of our approach. In Section 3 we describe the problem to be targeted and motivate its importance. Then, in Section 4 we present our contributions whose benefits are demonstrated in Section 5 via simulations. Finally, in Section 6 we draw conclusions and discuss possible future research directions.

2. Notation and Preliminaries

The system dynamics is modeled through the conservation of vehicles for both the internal and external queues. The first state equation gives the time evolution of the number of vehicles in the protected/controlled network (representing the evolution of “internal queues”) over a sample step of duration T , that is,

$$N_{k+1} = \left[N_k + T \left(\sum_{i \in \mathcal{I}} q_k^{\text{in},i} - \sum_{j \in \mathcal{O}} q_k^{\text{out},j} + \sum_{h \in \mathcal{D}} d_k^h \right) \right]^+, \quad (1)$$

where $[\cdot]^+$ is the maximum between zero and its argument, N_k denotes the number of vehicles in the network at discrete time-step k , $q_k^{\text{in},i}$ and $q_k^{\text{out},j}$ denote the inflow at link i and outflow at link j at time-step k in unit [veh/h], respectively. \mathcal{I} denotes the set of entrance queues and \mathcal{O} denotes the set of exit links. Let \mathcal{D} denote the set of entrance gates that are ungated. Then, d_k^h is the ungated but measured inflow from the entrance set \mathcal{D} , which cannot be compensated by the gated entry links.

Let $q_k^{\text{in}} \triangleq \sum_{i \in \mathcal{I}} q_k^{\text{in},i}$ and $q_k^{\text{out}} \triangleq \sum_{j \in \mathcal{O}} q_k^{\text{out},j}$, $d_k \triangleq \sum_{h \in \mathcal{D}} d_k^h$ equation (1) can be abstracted to a single internal queue, i.e.,

$$N_{k+1} = \left[N_k + T d_k + T (q_k^{\text{in}} - q_k^{\text{out}}) \right]^+. \quad (2)$$

The network outflow is modelled through the NFD concept, giving overall network flow Q as a concave function of network accumulation N . The total regional circulating flow $Q(N)$ is approximated by Edie’s generalized

definition of flow, i.e., the weighted average of link flows multiplied with link lengths. If we assume that the average trip length Υ in the network is constant and the average link length is given by l , then the output (throughput) of the network can be expressed as follows Daganzo (2007):

$$q_k^{\text{out}} = \frac{l}{\Upsilon} Q(N_k). \quad (3)$$

Output flow q_k^{out} is the estimated rate at which vehicles complete trips per unit time either because they finish their trip within the network or because they move outside the network. This function describes steady-state behavior of single-region homogeneous networks if the input to output dynamics are not instantaneous and any delays are comparable with the average travel time across the region Kulcsar et al. (2015).

Assumption 1. *The function $Q(N)$ is continuously differentiable and concave NFD over the eventual interval on N and network flow is uniform.*

Network inflow q_k^{in} is considered to be the controlled input of the system that follows the admission control policy. This flow depends on the external queue state, network state, performance requirements, and the network NFD. The admittance into the network is described through a simple queuing model, for entrance gate i , by:

$$L_{k+1}^i = \left[L_k^i + T \left(\lambda_k^i - q_k^{\text{in},i} \right) \right]^+, \quad (4)$$

where L_k^i is the queue length of the i th external queue and λ_k^i denotes the uncontrolled arrival rate at time k . We assume the arrival rate is an unknown, deterministic and bounded demand sequence. Exploiting that no negative queues may appear, by summing all external queues $i \in \mathcal{I}$,

$$L_{k+1} = \left[L_k + T \left(\lambda_k - q_k^{\text{in}} \right) \right], \quad (5)$$

where $L_k = \sum_{i \in \mathcal{I}} L_k^i$ and $\lambda_k = \sum_{i \in \mathcal{I}} \lambda_k^i$. Regarding the overall system, d_k and λ_k are considered as measured disturbance.

3. Problem statement

Similar to the classic perimeter control problem, the objective is to optimize network performance through the maximization of network through-

put. Moreover, the network performance is characterized by the performance requirements set. These performance requirements are usually specified for stochastic variables, e.g., the expected value of time delay, blockage probability of external queues. In this work, however, performance indicators are handled as deterministic values. By specifying upper/lower bounds for the indicators, hard constraints can be given for the system. For the traffic networks, two performance requirements are considered: (i) the average time delay in network should be less than a given threshold and (ii) the blockage of external queues should be avoided.

3.1. Average time delay in network

Suppose the network involves M distinct links. The average time delay is modeled by the following formula:

$$\Delta(N_k) = \frac{l}{v(N_k)} - \frac{l}{v_{\text{free}}}, \quad (6)$$

where l denotes the average link length of the network ($l = M^{-1} \sum_{i=1}^M l_i$) for links $i \in 1, \dots, M$, while $v(N_k)$ and v_{free} denote the actual and free link travel speed of the network, respectively.

According to Mahmassani et al. (1987), the generalized network-wide traffic flow variables, based on the extended Edie's definitions, can be expressed as follows:

$$\text{TTD}(N) = Q(N) \cdot T \sum_{i=1}^M l_i, \quad (7a)$$

$$\text{TTS}(N) = N \cdot T \sum_{i=1}^M l_i, \quad (7b)$$

$$v(N) = \frac{\text{TTD}(N)}{\text{TTS}(N)}, \quad (7c)$$

where $\text{TTD}(N)$ and $\text{TTS}(N)$ denote the Total Travel Distance and Total Time Spent in the network, respectively and average network speed is expressed as the quotient of these two. Substituting eqs. (7a) and (7b) into

(7c), the average network speed can be expressed as follows:

$$v(N_k) = \frac{Q(N_k)}{N_k}. \quad (8)$$

Note that $Q(N_k)$ is chosen such that $v(N_k)$ is an invertible function. In fact, it is intuitive that as the number of vehicles in the network N_k increases, the average speed of the network is expected to decrease. In virtue of Assumption 1, invertibility of $v(N_k)$ is therefore a direct consequence. The free travel speed can be approximated by the following formula:

$$v_{\text{free}} = \lim_{N_k \rightarrow 0^+} \frac{Q(N_k)^{(a)}}{N_k} \stackrel{(a)}{=} \lim_{N_k \rightarrow 0^+} \frac{\partial Q(N_k)}{\partial N_k}, \quad (9)$$

where (a) is due to L'Hôpital's rule.

Let τ_{free} denote the nominal travel time in the network when a vehicle travels with v_{free} and it is equal to l/v_{free} . We require that the average time delay in the network is smaller than a threshold value, herein denoted by Δ_{tr} , i.e., $\Delta(N_{k+1}) \leq \Delta_{\text{tr}}$.

3.2. Blockage of external queues

A deterministic approach is followed in which the aim is to avoid queue blockage, i.e., $L_k^i \leq L_{\text{cap},i}$ needs to be satisfied for all k and i , where $L_{\text{cap},i}$ denotes the capacity (maximally allowed queue length) of the i th external queue. This indicator is motivated by the need to avoid gridlocks in the external network through blocking the waiting queues.

Remark 1. *Arrival rates λ_i , $i \in \mathcal{I}$ are supposed to appear with a similar rate at each entrance link, therefore queues of similar length are built. In the control problem, the sum of all capacities of all external queues are used as a constraint.*

Remark 2. *Note that there may occur such high λ_k rates for which it is not possible to guarantee that both performance requirements are fulfilled.*

4. Main results

In this section, the problem is first cast as an optimization problem. Then, after algebraic manipulations, we restate our constraints (performance requirements) as upper and lower bounds of the internal queue length.

4.1. Performance conditions

The control aim is to maximize the network outflow (3) such that the specified performance conditions are satisfied. The outlined performance conditions can be formalized as follows:

- For the time delay, $\Delta(N_{k+1}) \leq \Delta_{\text{tr}}$ is given. This condition is used to guarantee a performance on the travel time vehicles spend in the region. It gives an upper bound for the internal accumulation N and thus the inflow to the region.
- External queue blockage is avoided if $L_{k+1} \leq L_{\text{cap}}$. In case of high arrival rates, prescription of this value leads to a lower bound for the internal queue, and indirectly for the inflow to the region.

Additionally, a constraint can be formalized for the admissible flow as follows:

$$0 \leq q_k^{\text{in}} \leq \min(\lambda_k + L_k/T, g_{\text{max}}s), \quad (10)$$

where g_{max} denotes the maximal green time of the entering links and it is equal to

$$g_{\text{max}} = \sum_{i \in \mathcal{I}} g_{\text{max},i},$$

where $g_{\text{max},i}$ denotes the maximal green time of input link i . The saturation flow of input links is assumed to be constant (for simplicity of exposition), and it is denoted by s . This constraint is not restricting the operation of the network and it basically states that the inflow cannot be less than zero or more than the amount of external queue which can be injected into the network. In the followings, the upper bound for vehicle inflow is denoted by $q_k^{\text{in,ub}} = \min(\lambda_k + L_k/T, g_{\text{max}}s)$.

The general optimization problem for an arbitrary horizon, say of size

m , can be cast as follows:

$$\max_{\{q_k^{\text{in}}, \dots, q_{k+m}^{\text{in}}\}} \sum_{\ell=1}^m Q(N_{k+\ell}) \quad (11a)$$

$$\text{subject to: } \Delta(N_{k+\ell}) \leq \Delta_{\text{tr}}, \quad \forall \ell \in 1, \dots, m \quad (11b)$$

$$\lambda_{k+\ell} = \lambda_k, \quad \forall \ell \in 1, \dots, m \quad (11c)$$

$$d_{k+\ell} = d_k, \quad \forall \ell \in 1, \dots, m \quad (11d)$$

$$L_{k+\ell} \leq L_{\text{cap}}, \quad \forall \ell \in 1, \dots, m \quad (11e)$$

$$0 \leq q_{k+\ell}^{\text{in}} \leq q^{\text{in,ub}} \quad \forall \ell \in 1, \dots, m \quad (11f)$$

$$N_{k+1} = [N_k + Td_k + T(q_k^{\text{in}} - q^{\text{out}}(N_k))]^+ \quad (11g)$$

$$L_{k+1} = [L_k + T(\lambda_k - q_k^{\text{in}})]. \quad (11h)$$

4.2. Relaxation of the optimal control problem

The receding horizon control problem in eq. (11) leads to a nonlinear optimization problem which is not convex. In the followings, the problem is reformalized as a single step control problem. The special connection between the external and internal dynamics gives basis for relaxing the problem to a convex optimization problem in which the only decision variable is the internal accumulation N_k .

The optimization problem can then be formulated as a one step ahead rolling one as

$$\max_{q_k^{\text{in}}} Q(N_{k+1}) \quad (12a)$$

$$\text{subject to: } \Delta(N_{k+1}) \leq \Delta_{\text{tr}} \quad (12b)$$

$$L_{k+1} \leq L_{\text{cap}} \quad (12c)$$

$$0 \leq q_k^{\text{in}} \leq q_k^{\text{in,ub}} \quad (12d)$$

$$N_{k+1} = [N_k + Td_k + T(q_k^{\text{in}} - q^{\text{out}}(N_k))]^+ \quad (12e)$$

$$L_{k+1} = [L_k + T(\lambda_k - q_k^{\text{in}})]. \quad (12f)$$

We hereby suggest the following optimal delay-aware traffic control policy.

Proposition 1. *Given a single-step control horizon with constraints (11b)-(12f) on state variables N_{k+1} and L_{k+1} and q_k^{in} , optimization problem (12)*

can be relaxed to a convex optimization problem:

$$\max_{N_{k+1}} Q(N_{k+1}) \quad (13a)$$

$$\text{subject to:} \quad N_{k+1}^{\text{lb}} \leq N_{k+1} \leq N_{k+1}^{\text{ub}} \quad (13b)$$

from which once (13) is solved, the optimal control input q_k^{in} can be calculated by (2). •

Proof 1. The upper and lower bounds are obtained as follows. By substituting the speed function $\Delta(N_k)$ from (6) into (11b), a constant lower bound can be derived for the speed, i.e.,

$$v^{\text{lb,delay}} = \frac{l}{\Delta_{\text{tr}} + \tau_{\text{free}}}. \quad (14)$$

The constant upper bound for the internal queue is obtained by inverting the speed function:

$$N^{\text{ub,delay}} = v^{-1} \left(\frac{l}{\Delta_{\text{tr}} + \tau_{\text{free}}} \right). \quad (15)$$

Substituting the upper bound for controlled inflow q_k^{in} from (12d) into the equality constraint (12e) a non-constant upper bound emerges and it is given by

$$N_{k+1}^{\text{ub,que}} = Tq_k^{\text{in,ub}} + N_k + Td_k - Tq^{\text{out}}(N_k). \quad (16)$$

As a result, the applied upper bound for the decision variable is given as the minimum of the upper bounds found in (15) and (16), i.e.,

$$N_{k+1}^{\text{ub}} = \min(N_k^{\text{ub,que}}, N^{\text{ub,delay}}). \quad (17)$$

Lower bound for N can be obtained by substituting (12e) and (12f) to (12c), i.e.,

$$N_{k+1}^{\text{lb,block}} = N_k + Td_k - Tq_k^{\text{out}}(N_k) + L_k + T\lambda_k - L_{\text{cap}}. \quad (18)$$

Note, that $N^{\text{lb,block}}$ may take negative values. Hence, the applied lower bound is given as:

$$N_{k+1}^{\text{lb}} = \max(0, N_k^{\text{lb,block}}). \quad (19)$$

Remark 3. Due to the min and max functions in the constraint description, we have nonlinear constraints that are usually simplified by a mixed integer formulation. In our approach, time-varying constraints are applied,

and assuming a polynomial NFD with a global maximum, the maximization of discharge flow leads to a convex optimization problem with new boundary constraints to be solved in each step. As a result, the mixed integer formulation is no longer needed. •

Remark 4. *Following Assumption 1, i.e. with concave MFDs, including the non-symmetric unimodal curve skewed to the right, the predictive control problem fits into the constrained convex optimization framework according to Proposition 1.*

Regarding the overall network that involves the external and internal queues, the performance requirements define a modified capacity of the system through the time varying interval of bounds. As noted in Remark 2, for a very large arrival rate λ_k it is not possible to guarantee that both performance requirements are fulfilled. This can be seen from (19), where as λ_k increases the lower bound becomes larger, and hence for large λ_k our lower bound may become larger than the upper bound. One of the main advantages of our method, is that it is able to detect when this situation occurs. In such situations, we need to prioritize between the performance conditions. In our scheme, priority is given to the vehicles in the protected network, i.e., violation of the upper bound, which corresponds to guaranteeing the average time delay in the network, is not permitted. Hence, when the lower bound becomes equal to or even exceeds the upper bound, at that time step the solution of the problem N_{k+1} is the upper bound itself, and no optimization is required to be solved.

Proposition 2. *The maximum arrival rate λ_k that can be handled by the network is found by restricting the lower bound of the vehicles in the network to be smaller than or equal to the upper bound, i.e., $N_{k+1}^{\text{lb}} \leq N_{k+1}^{\text{ub}}$. Thus,*

$$\lambda_k \leq \lambda_k^{\text{max}} \triangleq \max(0, N_{k+1}^{\text{ub}} + L_{\text{cap}} - N_k - Td_k + Tq_k^{\text{out}}(N_k) - L_k).$$

For any value above λ_k^{max} , by choosing N_{k+1} to be the solution to the optimization, we relax the constraint of having $L_{k+1} \leq L_{\text{cap}}$ for the external queues in order to keep the network flow at its maximum and avoid compromising the travel delay in the network. •

Proof 2. *The proof directly follows from Proposition 1.*

Remark 5. *In Proposition 2, N_{k+1}^{ub} compresses information on maximum green time and eventual changes in saturation flow.*

5. Simulation analysis

In the sequel, the previously proposed macroscopic admission control policy is tested over an emulated traffic network using a microscopic traffic simulator. The case study aims at comparing the following perimeter gating approaches, (i) a simple PI gating policy (see Appendix A.2), (ii) an MPC controller as described in Appendix A.3 and (iii) the proposed relaxed optimal scheme together (referred to as ‘relaxed controller’) with the (iv) fixed time strategy (uncontrolled case).

5.1. Simulation environment

For the simulations, the microscopic traffic simulator, Vissim (Fellendorf (1994)), is utilized. Through the COM interface, Vissim is connected to MATLAB, which is used for the online optimization of perimeter signal control. In each cycle, traffic measurements of the states are updated in MATLAB and new control signals are returned to the traffic simulator.

The signal and measurement update cycle of the network are equally 60s. Lengths of external queues are obtained by link measurements. Network accumulation and network average speed are calculated by aggregating individual link data. Network outflow is obtained as the sum of exit flows ($q_k^{\text{out}} \triangleq \sum_{j \in \mathcal{O}} q_k^{\text{out},j}$). The control input is computed as the overall network inflow, q_k^{in} , which is then divided to $q_k^{\text{in},i}$ entrance flows by following a rule detailed in Appendix A.4.

We expect the three controllers to show significantly different behaviours. The PI controller, being unaware of state constraints will aim to maximize outflow purely (resulting good travel times in unsaturated external network cases). However, in case of large demands a balanced loading of the external and internal networks is needed which is expected from the MPC and the relaxed controllers. When the constraints conflict, however, the MPC will not provide a clear policy on prioritizing, while the relaxed controller will follow Proposition 2.

5.2. Network model

The test network models the city center of Stockholm, Sweden (see Fig. 1).

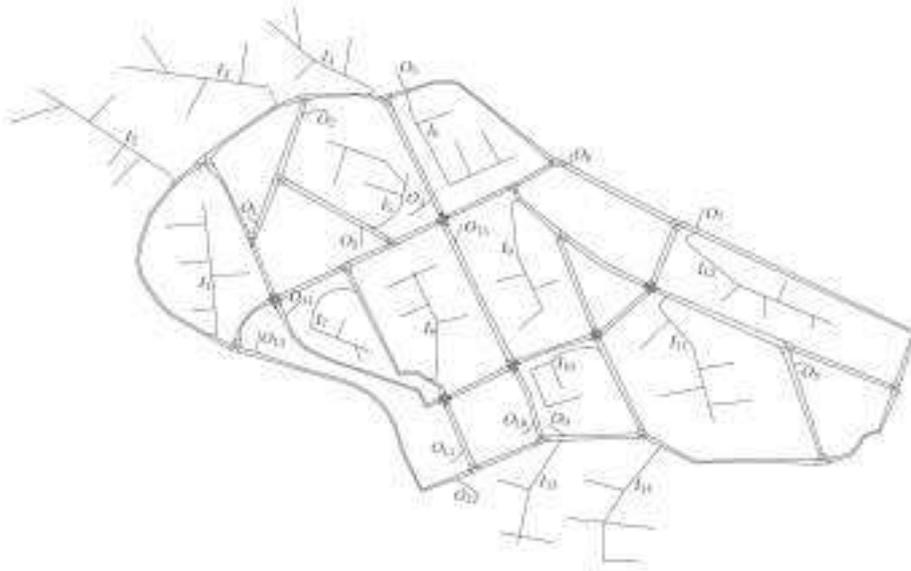


Figure 1: Layout of the Stockholm network, with entrance links $\{I_1, \dots, I_{14}\}$ and exit links $\{O_1, \dots, O_{15}\}$

Traffic enters the network through 14 entrance links having different link capacities, $L_{\text{cap},i}$. To the entrance links short links are connected through simple, priority-rule based intersections. Thereby the effect of the gating policy can be analyzed on the exterior network (e.g. through the occurrence of spillbacks). Network output is served by 15 exit links. In the simulations, fixed routing schemes are used among the origin-destination pairs. Inside the network, a fixed-time signal control is run at all the 24 intersections. Measurements are taken along 78 separate links, measuring average speed and the number of vehicles. The lengths of the longest and the shortest links are approximately 1.98 km and 0.33 km.

For the approximation of the NFD, a third-order form $Q(N_k) = p_3 N_k^3 + p_2 N_k^2 + p_1 N_k + p_0$ is applied (see Fig. 2), being concave on the domain $[0 N_{\text{max}}]$. Parameters p_3 to p_0 , with further model parameters are given in Appendix A.1, Table 2.

5.3. Case study

In the case study a three-hour-long rush hour scenario is analyzed featuring all three controllers. Network load and initial conditions are chosen such that the conflict of the two performance criteria can be analyzed. The

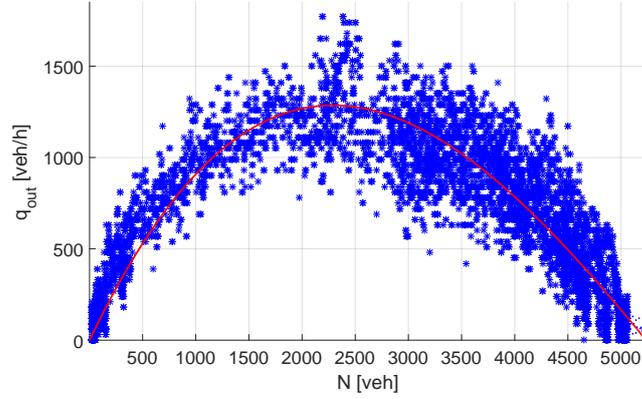


Figure 2: Network fundamental diagram of the simulation model

simulation results are plotted in Figs. 3-6. Fig. 3 depicts the arrival rate and the entrance flows of the different control situations. A sinusoid arrival rate is simulated.

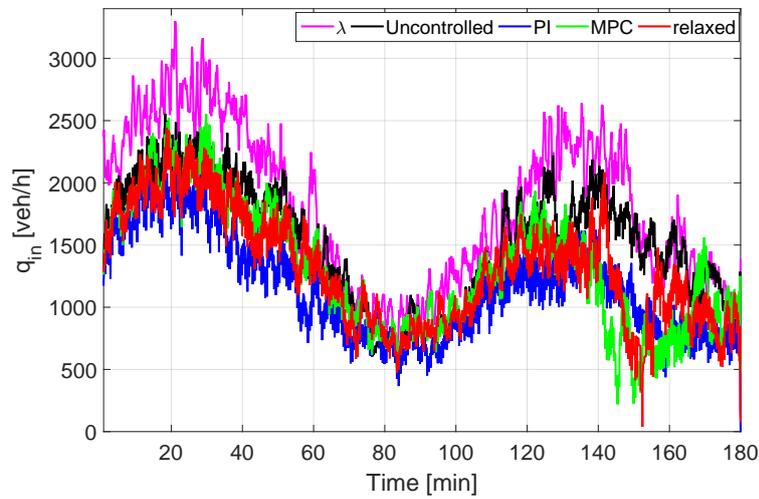


Figure 3: Arrival rate λ and controlled inflows

In case of no control, the network gets congested around 120min (see Fig. 4). The three controllers (PI, MPC, bluerelaxed) however manage to avoid congestions in the protected network.

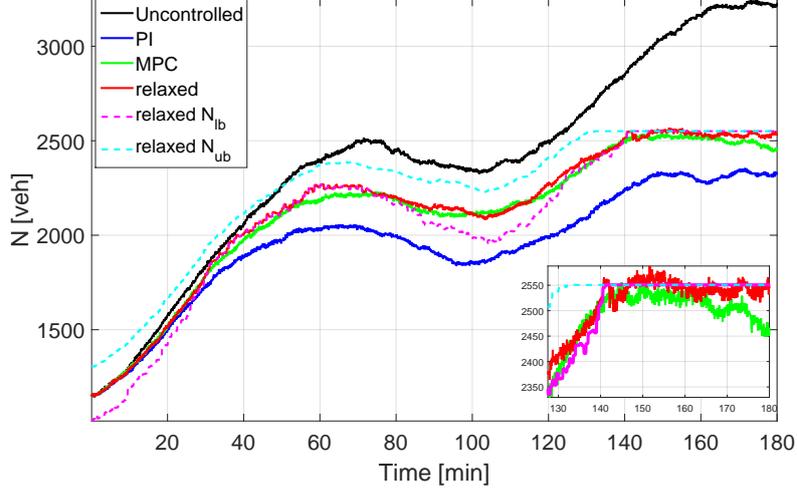


Figure 4: Number of vehicles in network with upper and lower bounds of the relaxed controller. During 128-180 min, state trajectories of the MPC and relaxed controllers are zoomed.

The PI control shows a fundamentally different behaviour from those of the MPC and relaxed controllers. It aims at tracking optimal network accumulation N_{opt} . This leads to a very good delay performance, however it entails the blockage of the external queues (Fig. 6). The reason for this is that the state bounds (11b) and (11e) cannot directly be applied to the PI controller. However, the bound for the input signal (11f) is satisfied due to the input saturation (21).

The MPC and relaxed approaches are very similar regarding the input signal. In the states, however, the different operation during the conflicting performance requirements can be observed. Starting at 130min, the constraint on travel delay (Fig. 5) and the external queues (Fig. 6) are approached due to the high arrival rates. This causes trouble for the MPC controller as it is not capable of satisfying both constraints, and therefore first violating the delay and then also the blockage constraint. Nevertheless, the relaxed controller is capable of prioritizing constraints, aiming to keep travel delay below Δ_{tr} and filling up and thus blocking external queues.

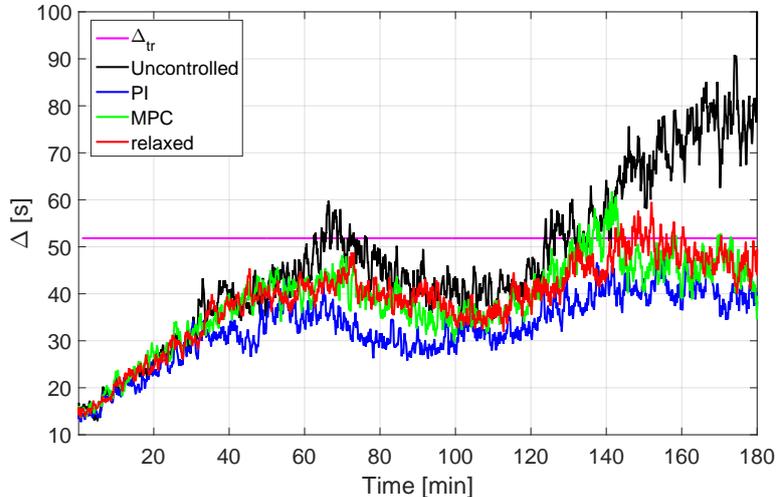


Figure 5: Average travel delay per link (calculated from network average speed). The allowed threshold value (see Appendix A.1) for time delay requires the traffic to circulate at least at 50% of the free speed.

The blockage of external queues results in a spillback outside the protected network. The total number of vehicles in the external network (queues) is plotted in Fig. 7. In case of link blockage, a sudden increase can be observed in the external vehicle number originating from spillback. In fact with L_{cap} the MPC and the relaxed admission control solutions indirectly address spillback. This is not the case with the uncontrolled nor with the PI controlled cases.¹

At certain points, the relaxed controller seems to violate the state constraint on the internal queue and as a result, the performance of travel delay (see the zoomed part of Fig.4 and Fig. 5). As shown in the zoomed plot, in this case, N_{lb} becomes larger than N_{ub} , and the former value needs to be followed as equality constraint. This is not completely satisfied, the relaxed controller tracks this value with some fluctuation. This is a result of the uncertainty in NFD modeling, as the outflow is not a deterministic function of the internal queue length.

¹Constraining L_{cap} , we indirectly account for spillback in the external links as long as the arrival rate is within the stable accommodation region.

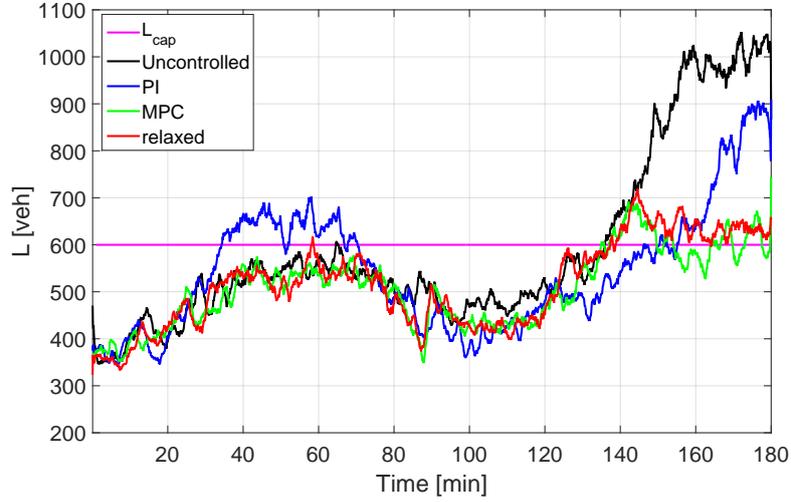


Figure 6: Sum of external queues

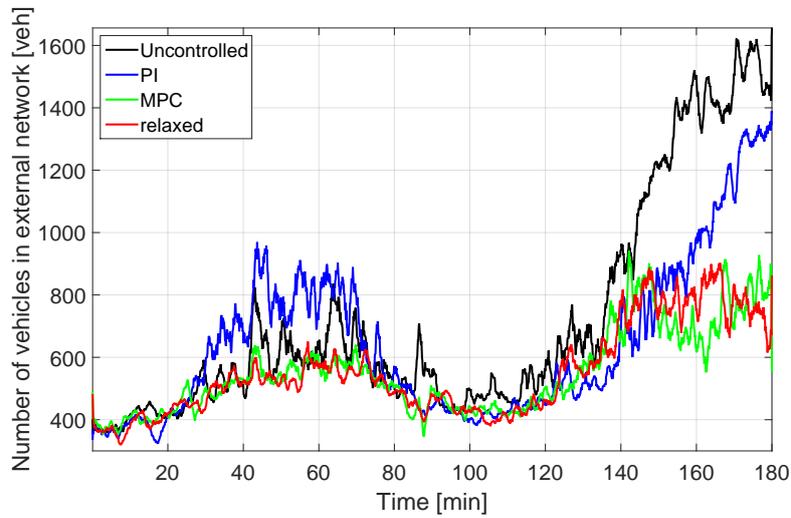


Figure 7: Number of vehicles in the external network

Regarding network outflow (Fig. 8), best performance is obtained by the PI controller, however, at the expense of blocking the external queues. There is no significant difference between the MPC and the relaxed controllers.

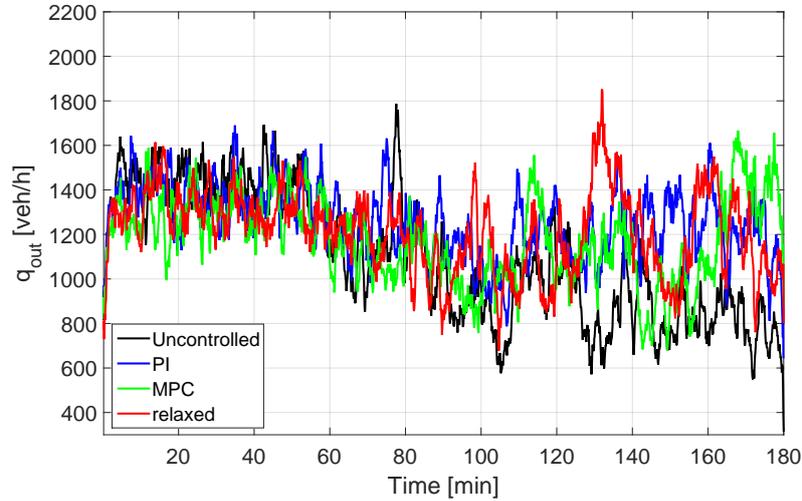


Figure 8: Network outflow

The overall delay performance for the external and internal network (i.e. analyzing a vehicle which travels from outside of the network through it to another external point) is represented by the overall TTS performance (see Fig. 9). Clearly, the PI controller provides the best performance until the saturation of the external network, from 160 min onwards. From that moment, the performance constrained controllers - the MPC and the relaxed approach - provide a balanced solution for mitigating the network-wide jam).

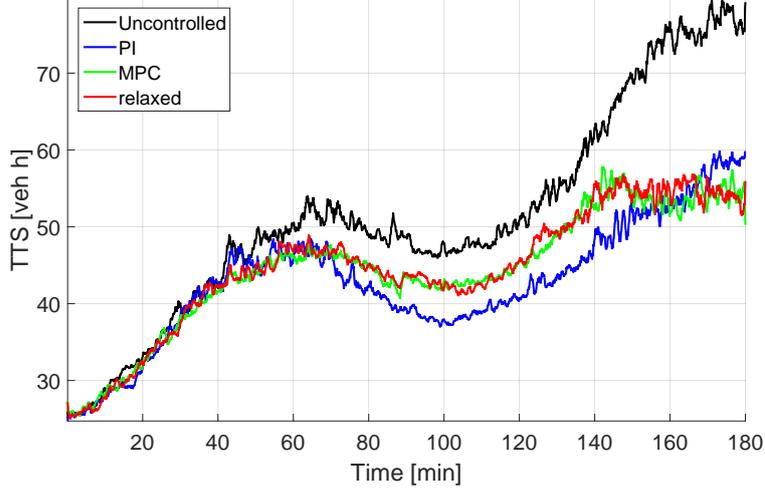


Figure 9: Overall TTS

The main advantage of constrained control solution can be summarized as follows.

- If $L_{cap} - L_k$ is large (i.e. the external queue buffer is far from being saturated), depending on N_k and $\Delta(N_k)$, the behaviour of the constrained controllers will be similar to that of the PI approach. They all control inflow such that the obtained accumulation of the internal network gives the best outflow. This case can be observed during 0 to 30 min and 70 to 130 min intervals: the three controllers show a similar rate for loading the network, while both constraints are fulfilled (Figs. 3, 4 and 6).
- In case $L_{cap} - L_k$ is small, and the internal network has room for accommodating vehicles, q_{in} will be as large as possible, similar to a no control case. However, delay constraint is enforced, which the PI control will not satisfy. This behaviour can be captured from 30 to 70 min in Fig 4: around 30 min the PI control starts to diverge from the accumulation (in Fig. 4) which violates the bounds on external blockage (given through eq. (18)). As Fig. 6 shows, queue capacity bound is not satisfied from 30 to 70 min by the PI solution whereas the MPC and the relaxed approach keeps both constraints.

- In case if $L_{cap} - L_k$ is small, and the internal network has high accumulation, for the constrained controllers, q_{in} will be gated to maximize capacity under internal delay bound. If constraints are conflicting, with relaxed admission controller, we prioritise the protected network and drop the external link capacity constraint as discussed in Proposition 2. The PI controller however would never respect external constraint directly. This case can be recognized from 120 min onwards in Fig. 4 and 6. The external queue reserves are low, and the internal accumulation is close to bounds. From 120 min the internal bound is active, but not violated until 140 min, where the lower bound meets the upper bound. Then, the rule of Proposition 2 is followed, prioritizing the internal accumulation bound to the external one. This is only satisfied by the relaxed controller, while the MPC has conflicting constraints, neither of which is satisfied.

Apart from the handling of performance constraints, the relaxed controller shows an appealing performance in computational time (see Table 1), due to the relaxation of the problem and the single-step control horizon.

Method	PI	MPC	relaxed
Comp. time [s]	$6 \cdot 10^{-4}$	0.241	0.009

Table 1: Computation time of a sample step. (Simulations are run on a PC with Intel i5 3.0GHz CPU and 8GB RAM.)

6. Conclusions and future directions

6.1. Conclusions

In this paper we proposed an admission control mechanism that maximizes network outflow while specified performance requirements are satisfied. These performance requirements were incorporated as constraints into the system.

First, a predictive, constrained optimization problem was formulated. Next, the problem was reformulated as a single-step convex optimization problem, and an algorithm was developed ensuring throughput maximization subjected to network travel time constraint guarantees. The performance of our approach was demonstrated via case studies and compared

to that of the PI control and MPC control approaches. The case studies illustrate that the proposed mechanism has improved performance in terms of network throughput, average time delay and external queue length.

6.2. Future Directions

Since the shape of the NFD is affected by different factors, it is important to study the problem under uncertain traffic flow description. Towards this end, Kulcsar et al. (2015) proposed an \mathcal{L}_2 optimal control design, and Haddad and Shraiber (2014) a robust control one, based on the Linear Parameter-Varying (LPV) model structure. However, none of these approaches incorporated performance requirements, which is part of our ongoing research.

In the case of heterogeneous networks, a set of homogeneous subregions can be defined, described by individual NFDs Ramezani et al. (2015). In this case, multi-region perimeter control is used. In Aboudolas and Geroliminis (2013) perimeter and boundary control is developed via multivariable Linear Quadratic (LQ) regulators. In Hajiahmadi et al. (2013) the problem of route guidance is solved for a multi-region network. Furthermore, Geroliminis et al. (2013) and Haddad et al. (2013) propose cooperative subregion controllers in a predictive control framework. As mentioned in the introduction, new MFD-based model for two-region networks with aggregate boundary queue dynamics is introduced in Haddad (2017b), where maximum lengths are imposed on the aggregate boundary queues, that aim at maintaining the existence of well-defined MFDs and their dynamics. Unlike our work, they cannot relax the maximum length of queues due to the fact that the limitation is based on the network structure and not on some performance target. Moreover, in Haddad (2017b) the outflow of one network does affect the inflow of the other. We do not consider such a coupling in our work. More recently, Haddad and Mirkin (2017) propose the use of adaptive control in order to deal with uncertainties and take into consideration the restrictions on the available information in the perimeter control problem for multi-region (several interconnected homogeneous regions) MFD systems. Current research focuses on extending this work to consider the admission control problem for multiple regions interacting with each other, where each region has as external queues, and hence, gates (part of) the internal queues of other regions, while at the same time, on contrary to existing work in the literature, certain performance requirements are guaranteed.

A. Appendix

A.1. Network parameters

Parameter	Value
p_3	$1.864 \cdot 10^{-6}$
p_2	-0.0003308
p_1	1.221
p_0	0
T	60 s
N_{opt}	2290 veh
N_{max}	5000 veh
l/Υ	0.0111
l	0.6047 km
s	0.5 veh/s/lane
$g_{\text{max},i}$	40s $\forall i \in \mathcal{I}$
v_{nom}	42 km/h
τ_{free}	51.8 s
Δ_{tr}	51.8 s
$L_{\text{cap},i}$	[35, 50, 40, 45, 40, 30, 55, 50, 60, 40, 30, 45, 40, 40] veh

Table 2: Model parameters

A.2. PI control

The control rule is similar to the one applied in Keyvan-Ekbatani et al. (2012):

$$q_k^{\text{in,PI}} = q_{k-1}^{\text{in}} + K_I(N_k - N_{k-1}) + K_P(N_{\text{opt}} - N_k), \quad (20)$$

where K_P and K_I are the control design parameters, obtained by manual tuning. The design resulted in the following values: $K_I=0.3$, $K_P=0.07$. Input saturation is applied for the controller in the following form:

$$q_k^{\text{in,sat}} = \min(q_k^{\text{in,ub}}, q_k^{\text{in,PI}}). \quad (21)$$

where $q_k^{\text{in,ub}}$ is given in eq. (10).

A.3. MPC control

MPC is well suited to this problem, since it is a direct constraint handling method that can be implemented over a finite prediction horizon. Optimization problem (12) is now adapted to MPC framework. First, an equality constraint is involved for the disturbance: throughout the control horizon, λ_k is considered constant. Furthermore, the decision variable of the optimization is the vehicle inflow q_k^{in} instead of the internal queue N_{k+1} , and distinctively, bounds are defined for the states and the control input. The cost function is also extended. The first term implies the optimization of discharge flow of the protected network. The demand matching in the second term is given to avoid an unnecessary suppression of inflow. The first two terms thus lead to a balanced control of the internal and external queues. The third term is applied to suppress input oscillations. Hence, the optimization problem for the MPC framework is given by

$$\min_{[q_k^{\text{in}}, \dots, q_{k+m}^{\text{in}}]} \sum_{\ell=1}^m \left\{ -Q(N_{k+\ell}) + \|q_{k+\ell}^{\text{in}} - \lambda_{k+\ell}\|_2^2 + \|q_{k+\ell}^{\text{in}} - q_{k+\ell-1}^{\text{in}}\|_2^2 \right\} \quad (22a)$$

$$\text{subject to: } N_{k+1} = N_k + Td_k + T[q_k^{\text{in}} - q^{\text{out}}(N_k)]^+ \quad (22b)$$

$$L_{k+1} = L_k + T[\lambda_k - q_k^{\text{in}}] \quad (22c)$$

$$\lambda_{k+\ell} = \lambda_k, \quad \forall \ell \in 1, \dots, m \quad (22d)$$

$$d_{k+\ell} = d_k, \quad \forall \ell \in 1, \dots, m \quad (22e)$$

$$0 \leq L_{k+\ell} \leq L_{\text{cap}} \quad \forall \ell \in 1, \dots, m \quad (22f)$$

$$0 \leq q_{k+\ell}^{\text{in}} \leq \min\left(\lambda_{k+\ell} + \frac{L_{k+\ell}}{T}, g_{\text{max}s}\right) \quad \forall \ell \in 1, \dots, m. \quad (22g)$$

The controller solves a convex optimization problem in a rolling horizon manner Grüne and Pannek (2011). For the case study, a control horizon of $N_c = 5$ applied.

A.4. Division of inflow to multiple entrances

Arrival rates λ_i , $i \in \mathcal{I}$ are supposed to appear with an equal rate at each entrance link, therefore queues of similar length are assumed to be built. In spite of this assumption, different queue lengths may be present due to an uneven load of the network. To maintain an equable load of external links, a simple rule is followed to divide input flows, detailed below.

First, define the capacity reserve of external queue i :

$$L_k^{\text{res},i} = \max(0, L_{\text{cap},i} - L_k^i). \quad (23)$$

Weighting factor α_i represents the proportion of capacity reserves:

$$\alpha_k^i = 1 - \frac{L_k^{\text{res},i}}{\sum_{i \in \mathcal{I}} L_k^{\text{res},i}}, \quad (24)$$

whereas β_i gives the fraction of all queued traffic at entrance i :

$$\beta_k^i = \frac{L_k^i}{\sum_{i \in \mathcal{I}} L_k^i}. \quad (25)$$

Overall weighting of inputs are given by μ_k^i as a combination of α_k^i and β_k^i :

$$\mu_k^i = \frac{\alpha_k^i \beta_k^i}{\sum_{i \in \mathcal{I}} \alpha_k^i \beta_k^i}. \quad (26)$$

Controlled inflow $q_k^{\text{in},i}$ at entrance i is then calculated as

$$q_k^{\text{in},i} = \mu_k^i q_k^{\text{in}} \quad (27)$$

As a result of the above rule, zero input is given to the entrances with queues of zero length; also, the highest input is given to the queues which are blocked or close to blocking. The rule is proportional to the degree of blockage, giving higher input priority to the links that have more waiting traffic beyond the blocked capacity.

Acknowledgements

This research is supported by Chalmers' initiatives in transport research, the Transport Area of Advance at Chalmers University of Technology and SAFER (Vehicle and Traffic Safety Centre) and by the National Research, Development and Innovation Office of Hungary - NKFIH through grant No. 115694. B. Kulcsar gratefully acknowledges the support received from Aeje.

References

Aboudolas, K., Geroliminis, N., 2013. Perimeter and boundary flow control in multi-reservoir heterogeneous networks. *Transportation Research Part B: Methodological* 55, 265–281.

- Ampountolas, K., Kouvelas, A., 2015. Real-time estimation of critical vehicle accumulation for maximum network throughput. In: American Control Conference (ACC). pp. 2057–2062.
- Bigazzi, A. Y., Figliozzi, M. A., 2012. Congestion and emissions mitigation: A comparison of capacity, demand, and vehicle based strategies. *Transportation Research Part D: Transport and Environment* 17 (7), 538–547.
- Csikós, A., Tettamanti, T., Varga, I., 2015. Nonlinear gating control for urban road traffic network using the network fundamental diagram. *Journal of Advanced Transportation* 49 (5), 597–615.
- Daganzo, C., Geroliminis, N., 2008. An analytical approximation for the macroscopic fundamental diagram of urban traffic. *Transportation Research Part B: Methodological* 42 (9), 771–781.
- Daganzo, C. F., 2007. Urban gridlock: Macroscopic modeling and mitigation approaches. *Transportation Research Part B: Methodological* 41 (1), 49–62.
- Fellendorf, M., 1994. Vissim: A microscopic simulation tool to evaluate actuated signal control including bus priority. In: 64th Institute of Transportation Engineers Annual Meeting. Springer, pp. 1–9.
- Geroliminis, N., Daganzo, C. F., 2008. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological* 42 (9), 759–770.
- Geroliminis, N., Haddad, J., Ramezani, M., March 2013. Optimal Perimeter Control for Two Urban Regions With Macroscopic Fundamental Diagrams: A Model Predictive Approach. *IEEE Transactions on Intelligent Transportation Systems* 14 (1), 348–359.
- Geroliminis, N., Sun, J., 2011. Properties of a well-defined macroscopic fundamental diagram for urban traffic. *Transportation Research Part B: Methodological* 45 (3), 605 – 617.
- Godfrey, J. W., 1969. The mechanism of a road network. *Traffic Engineering and Control* 11 (7), 323–327.
- Grüne, L., Pannek, J., 2011. *Nonlinear Model Predictive Control: Theory and Algorithms*. Springer London, London.
- Haddad, J., 2017a. Optimal coupled and decoupled perimeter control in one-region cities. *Control Engineering Practice* 61, 134 – 148.
- Haddad, J., 2017b. Optimal perimeter control synthesis for two urban regions with aggregate boundary queue dynamics. *Transportation Research Part B: Methodological* 96, 1 – 25.
- Haddad, J., Mirkin, B., 2016. Adaptive perimeter traffic control of urban road networks based on MFD model with time delays. *International Journal of Robust and Nonlinear Control* 26, 1267–128.
- Haddad, J., Mirkin, B., 2017. Coordinated distributed adaptive perimeter control for large-scale urban road networks. *Transportation Research Part C: Emerging Technologies* 77, 495 – 515.
- Haddad, J., Ramezani, M. G., Geroliminis, N., 2013. Cooperative traffic control of a mixed network with two urban regions and a freeway. *Transportation Research Part B: Methodological* 54, 17–36.
- Haddad, J., Shraiber, A., 2014. Robust perimeter control design for an urban region. *Transportation Research Part B: Methodological* 68, 315–332.

- Hajiahmadi, M., Haddad, J., Schutter, B. D., Geroliminis, N., March 2015. Optimal Hybrid Perimeter and Switching Plans Control for Urban Traffic Networks. *IEEE Transactions on Control Systems Technology* 23 (2), 464–478.
- Hajiahmadi, M., Knoop, V., Schutter, B. D., Hellendoorn, H., 2013. Optimal dynamic route guidance: a model predictive approach using the macroscopic fundamental diagram. In: *Proceedings of the 16th International IEEE Annual Conference on Intelligent Transportation Systems (ITSC)*. pp. 1022–1028.
- Helbing, D., 2009. Derivation of a fundamental diagram for urban traffic flow. *The European Physical Journal B* 70 (2), 229–241.
- Jenelius, E., Koutsopoulos, H. N., 2013. Travel time estimation for urban road networks using low frequency probe vehicle data. *Transportation Research Part B: Methodological* 53, 64 – 81.
- Keyvan-Ekbatani, M., Kouvelas, A., Papamichail, I., Papageorgiou, M., 2012. Exploiting the fundamental diagram of urban networks for feedback-based gating. *Transportation Research Part B: Methodological* 46 (10), 1393–1403.
- Klessig, H., Fettweis, G., 2014. Adaptive Admission Control in Interference-Coupled Wireless Data Networks: A Planning and Optimization Tool Set. In: *Proceedings of the IEEE International Conference on Communications (ICC) - Mobile and Wireless Networking Symposium*. pp. 2375–2380.
- Kouvelas, A., Saeedmanesh, M., Geroliminis, N., 2017. Enhancing model-based feedback perimeter control with data-driven online adaptive optimization. *Transportation Research Part B: Methodological* 96, 26 – 45.
- Kulcsar, B., Ampountolas, K., Dabiri, A., July 2015. Single-region robust perimeter traffic flow control. In: *European Control Conference (ECC)*. pp. 2628–2633.
- Le, L. B., Modiano, E., Shroff, N., Aug 2012. Optimal Control of Wireless Networks With Finite Buffers. *IEEE/ACM Transactions on Networking* 20 (4), 1316–1329.
- Leclercq, L., Chiabaut, N., Trinquier, B., 2014. Macroscopic fundamental diagrams: A cross-comparison of estimation methods. *Transportation Research Part B: Methodological* 62, 1 – 12.
- Li, R., Rose, G., 2011. Incorporating uncertainty into short-term travel time predictions. *Transportation Research Part C: Emerging Technologies* 19 (6), 1006–1018.
- Lin, S., Schutter, B. D., Xi, Y., Hellendoorn, H., 2012. Efficient network-wide model-based predictive control for urban traffic networks. *Transportation Research Part C: Emerging Technologies* 24, 122 – 140.
- Liu, C., Leung, K., Gkelias, A., 2014. A Generic Admission-Control Methodology for Packet Networks. *IEEE Transactions on Wireless Communications* 13 (2), 604–617.
- Mahmassani, H., Williams, J., Herman, R., 1987. Performance of urban traffic networks. In: *10th International Symposium on Transportation and Traffic Theory*. Amsterdam, The Netherlands. pp. 1–20.
- Mahmassani, H. S., Saberi, M., Zockaie, A., 2013. Urban network gridlock: Theory, characteristics, and dynamics. *Transportation Research Part C: Emerging Technologies* 36, 480 – 497.
- Papageorgiou, M., Diakaki, C., Dinopoulou, V., Kotsialos, A., Wang, Y., 2003. Review of road traffic control strategies. *Proceedings of the IEEE* 91 (12), 2043 – 2067.
- Rahmani, M., Jenelius, E., Koutsopoulos, H. N., 2015. Non-parametric estimation of route travel time distributions from low-frequency floating car data. *Transportation*

- Research Part C: Emerging Technologies 58, Part B, 343 – 362.
- Ramezani, M., Haddad, J., Geroliminis, N., 2015. Dynamics of heterogeneity in urban networks: aggregated traffic modeling and hierarchical control. *Transportation Research Part B: Methodological* 74, 1–19.
- Vlahogianni, E. I., Karlaftis, M. G., Golias, J. C., 2014. Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies* 43, Part 1, 3 – 19.
- Yildirimoglu, M., Geroliminis, N., 2014. Approximating dynamic equilibrium conditions with macroscopic fundamental diagrams. *Transportation Research Part B: Methodological* 70, 186 – 200.
- Yildirimoglu, M., Ramezani, M., Geroliminis, N., 2015. Equilibrium analysis and route guidance in large-scale networks with MFD dynamics. *Transportation Research Part C: Emerging Technologies* 59, 404 – 420.
- Zhan, X., Hasan, S., Ukkusuri, S. V., Kamga, C., 2013. Urban link travel time estimation using large-scale taxi data with partial information. *Transportation Research Part C: Emerging Technologies* 33, 37 – 49.