Comparison of On-Line Time-Delayed and Non-Time-Delayed Urban Traffic Control via Remote Gating

Dr. Mehdi Keyvan-Ekbatani*
Post-Doctoral Researcher
Department of Transport and Planning
Faculty of Civil Engineering and Geosciences
Delft University of Technology
Delft, The Netherlands
Phone: +31 (15) 27 86302
Email: m.ekbatani@tudelft.nl

Prof. Markos Papageorgiou
Professor, Director
Dynamic Systems and Simulation Laboratory
School of Production Engineering and Management
Technical University of Crete
73100 Chania, Greece
Phone: +30-28210-37289
Fax: +30-28210-37584
Email: markos@dssl.tuc.gr

Dr. Victor. L. Knoop
Assistant Professor
Department of Transport and Planning
Faculty of Civil Engineering and Geosciences
Delft University of Technology
Delft, The Netherlands
Phone: +31 (15) 278 8413
Email: v.l.knoop@tudelft.nl

***Corresponding Author**
Number of words: 5750 + (5 FIGURES) 1250 + (2 Table) 500 = 7500
Submitted only for Presentation at the 2015 Transportation Research Board 94th Annual Meeting
November, 2014

ABSTRACT
Recent studies demonstrated the efficiency of feedback-based gating control in mitigating congestion in urban networks by exploiting the notion of network fundamental diagram (NFD). The employed feedback regulator of proportional-integral(PI)-type targets an operating NFD point of maximum throughput to enhance the mobility in the urban road network during the peak period, under saturated traffic conditions. In previous studies, gating was applied directly at the border of the protected network (PN), i.e. the network part to be protected from over-saturation. In this work, the recently developed feedback-based gating concept is applied at junctions located further upstream of the PN. This induces a time-delay, which corresponds to the travel time needed for gated vehicles to approach the PN. The resulting extended feedback control problem can be also tackled by use of a PI-type regulator, albeit with different gain values compared to the case without time-delay. Detailed procedures regarding the appropriate design of related feedback regulators are provided. In addition, the developed feedback concept is shown to work properly with very long time-steps as well. A large part of the Chania, Greece, urban network, modelled in a microscopic simulation environment under realistic traffic conditions, is used as test-bed in this study. The reported results demonstrate a stable and efficient behaviour and improved mobility of the overall network in terms of mean speed and travel time.

Keywords: Over-saturated urban networks; network fundamental diagram (NFD); time-delayed feedback gating control

1 INTRODUCTION

Many different real-time Urban Traffic Control (UTC) strategies have been proposed up to now, but due to the reduced efficiency of the strategies (e.g. SCOOT [1] and SCATS [2]) in saturated traffic conditions, there is still space for further developments. Some other recently proposed UTC concepts employ computationally expensive algorithms, which might render the network-wide implementation of these optimization-based approaches difficult in terms of real-time feasibility; see e.g. [3]. [4] used a bang-bang like concept for control of an urban network with oversaturated intersections. However, these methods may allow the traffic to enter into the network part to be protected from over-saturation and act after the congestion starts occurring; hence, the methods should be specifically tested for highly congested urban road network parts.

Gating or metering is a practical tool, frequently employed against over-saturation of significant or sensitive links, arterials or urban network parts [5]. The idea is to hold traffic back (via prolonged red phases at traffic signals) upstream of the links to be protected from over-saturation, whereby the level or duration of gating may depend on real-time measurements from the protected links. Gating is usually employed in an ad hoc way (based on engineering judgment) regarding the specific gating policy and quantitative details, which may lead to insufficient or unnecessarily strong gating actions.

Recently, the reproducible relationship between flow and density occurring at the network level under certain conditions (e.g. homogeneous spatial distribution of the congestion), known as network fundamental diagram (NFD), has gained increased popularity. The notion had been initially proposed by [6]; but the verification of its existence with dynamic features is recent [7]. Although the NFD notion is still under investigation in various aspects, it can be exploited as a fruitful basis for derivation of urban signal control approaches. In particular, NFD is useful to introduce elegant perimeter traffic signal control strategies, as details of individual links are not needed to describe the congestion level and its evolution. Other interesting studies related to network level relations for simulated and real networks can be found in ([8], [9], [10], [11]) and others.

[12] used the NFD concept to propose a control rule that maximizes the network outflow; however, the proposed rule cannot be directly employed for practical use in urban networks. Based on the NFD, [13] developed a bang-bang control strategy to optimize the accumulated vehicular number. [14] pursued a model-predictive control (MPC) approach. However, MPC calls for sufficiently accurate model and external disturbance predictions, which may be a serious impediment for practicable control. [15] introduced a fixed-time signal timing perimeter control by exploiting the NFD, albeit without adaptation to the prevailing real-time traffic conditions. [16] proposed a generic real-time feedback-based gating concept, which exploits the urban NFD for smooth and efficient traffic control operations, with an application to the network of Chania, Greece. [17] exploited the NFD notion for route guidance. [18] demonstrated that efficient feedback-based gating is actually possible with much less real-time measurements, i.e. at much lower implementation cost, than in full network instrumentation. Given that in large metropolitan urban networks the congestion spreads mostly heterogeneously over the network, in which case a homogenous gating strategy may not be the optimal solution, [19] proposed a multiple concentric-boundary gating strategy, which implements the aforementioned feedback-based gating strategy, along with considering the heterogeneity of a large-scale urban network. [20] extended the basic feedback approach for application to multiple sub-networks with separate individual NFDs in a heterogeneous urban network. [21] introduced a three-dimensional NFD relating the accumulation of cars and buses with the total circulating flow in the network. They addressed that this finding may be applied for a perimeter flow control concept which maximizes the network capacity of vehicles or the passenger capacity in bi-modal urban networks. [22] proposed a proportional controller for a perimeter control strategy by exploiting the 3D-NFD concept. This control strategy might be beneficial by combining the 3D-NFD with the idea of reduced NFD (i.e. using few real-time measurements) which is thoroughly discussed in [18].

In all of the aforementioned gating strategies, metering is applied at the border of the protected network (PN). Based on restrictions which might be faced for implementing gating control at the perimeter of the PN in different urban networks, such as unavailability of proper links to store the gated vehicles (queuing) or sufficient number of signalized junctions, the traffic flow bound for the PN may be also metered at some junctions remote from the border of the PN. In this work, we apply the feedback gating strategy further upstream of the PN and design a robust feedback controller by considering a time-delay term, which corresponds to the travel time needed for gated vehicles to approach the PN (when the gating link is not directly at the PN perimeter). Detailed procedures regarding the appropriate design of related feedback regulators are provided. Moreover, the developed

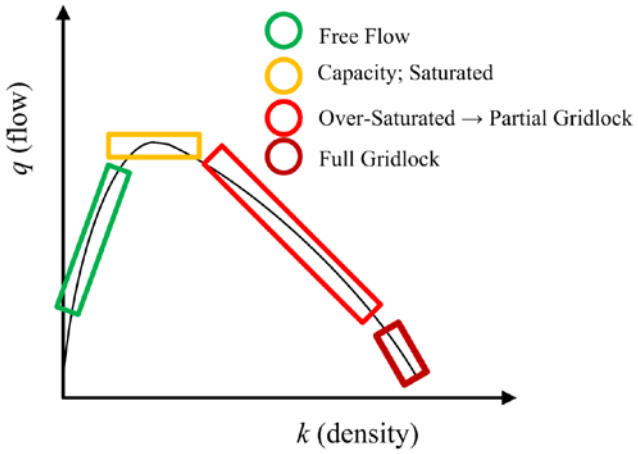


Figure 1 Network Fundamental Diagram (network total weighted flow vs. average density)

feedback concept is shown to work properly with very long time-steps as well. A large part of the Chania, Greece, urban network, modeled in a microscopic simulation environment under realistic traffic conditions, is used as test-bed in this study.

2 A BRIEF REVIEW OF NFD AND GATING CONCEPT

2.1 Network fundamental diagram

The reproducible relationship between the number of vehicles inside the network (or accumulation or average density) on one hand, and the link-length weighted total traffic flow of the network on the other hand, forms an asymmetric inverse-U shaped diagram (see Fig. 1) that is often called macroscopic fundamental diagram (MFD) [7]. For a better distinction of the MFD of highways from the (fundamentally different) MFD of an urban network, we call the latter a network fundamental diagram (NFD). According to the definitions introduced in [16], a network fundamental diagram is; (a) an *ideal* NFD, if based on exact knowledge of the displayed quantities (this is practically only possible in analytic or simulation-based studies) for all links $z \in \mathbb{Z}$, where $\mathbb{Z}$ is the set of all network links; or (b) an *operational* NFD, if based on available (more or less accurate) measurements and estimates at a subset $\mathbb{M}$ of all links, i.e. $\mathbb{M} \subseteq \mathbb{Z}$

In this paper we will only deal with operational NFD. The NFD's y-axis reflects the Total Travelled Distance (*TTD* in veh·km per h), while the x-axis reflects the Total Time Spent (*TTS* in veh·h per h) by all monitored vehicles in the PN. The *TTD* and *TTS* are obtained from the (emulated) loop measurements via the following equations:

$$TTS(k) = \sum_{z \in \mathbb{M}} \frac{T \cdot \hat{N}_z(k)}{T} = \sum_{z \in \mathbb{M}} \hat{N}_z(k) = \hat{N}(k) \tag{1}$$

$$TTD(k) = \sum_{z \in \mathbb{M}} \frac{T \cdot q_z(k) \cdot L_z}{T} = \sum_{z \in \mathbb{M}} q_z(k) \cdot L_z \tag{2}$$

where *z* is the link where a measurement is collected; $\mathbb{M}$ is the set of measurement links; $k = 0,1,2,\ldots$ is a discrete time index reflecting corresponding cycles; T is the cycle time; q_z is the measured flow in link *z* during cycle *k;* L_z is the length of link *z*; and $\hat{N}_z(k)$ is the estimated number of vehicles in link *z* during cycle *k*, which is derived from occupancy measurements via the following equation

$$\hat{N}_z(k) = L_z \cdot \frac{\mu_z}{100\lambda} \cdot o_z(k-1) \tag{3}$$

where $o_z(k)$ is the measured time-occupancy (in %) in link z during cycle k; μ_z is the number of lanes of link z; and λ is the average vehicle length. (3) is reasonably accurate, particularly if the detector is located around the middle of the link [23].

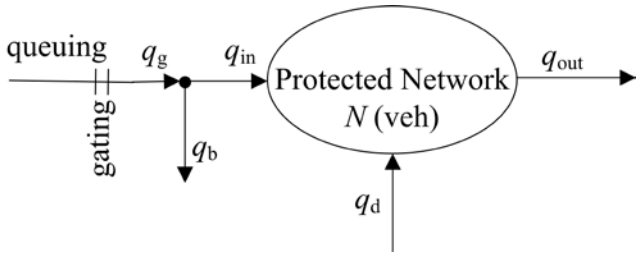
2.2 Gating concept

The general scheme of gating, including the protected network (PN), is sketched in Fig. 2(a). To implement gating, the usual traffic lights settings must be modified at (one or more) upstream junctions, which may be located at the border or a bit further of the problematic area. The gating location is shown by double line, upstream of which vehicle queues may grow temporarily faster than without gating; q_g is the gated flow, a part of which (q_b) may not be bound for the protected network (PN); while q_{in} is the part of the gated flow that enters the protected network; q_d represents other (non-gated or internal) inflows to the PN (disturbances); finally q_{out} and N stand for the PN exit flow (both internal and external) and the number of vehicles included in the PN, respectively.

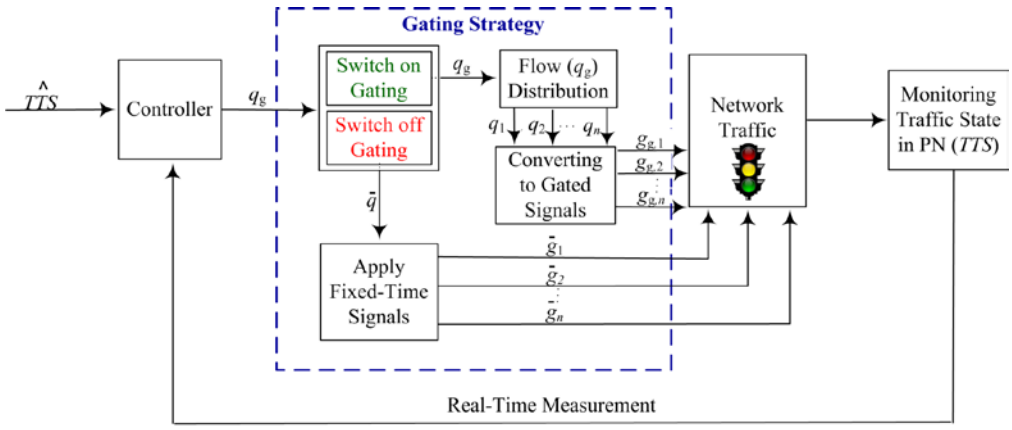
2.3 General gating procedure at the gated junctions

According to Fig. 2(b), the gating procedure at the gated signalized junctions can be listed as follows:

- The controller runs in background all the time; is fed with real-time measurements of PN (*TTS*), averaged over the last signal cycle; targets a set-point $\hat{TTS}$ (selected to be inside the critical range of NFD, see the yellow zone in Fig. 1) for maximum throughput; and produces an ordered gated flow q_g to be applied in the next signal cycle.
- A switch-on/off logic decides for the implementation of the ordered flow q_g by the controller; Typically, gating is applied only if the measured *TTS* is in a neighborhood of the set-point $\hat{TTS}$.
- During the switch-on period, the ordered flow q_g is distributed among the gated junctions ($q_1,...,q_n$) where n is the number of gated junctions.



(a) General scheme of gating (from [16])



(b) General gating procedure of a gated junction

FIGURE 2 Gating traffic control

- The distributed flows are converted into green phases ($g_{g,1},...,g_{g,n}$) where $g_{g,i}$, i=1,…,n, are the gated green times.
- During the switch-off period, the ordinary fixed-time plan is implemented ($\bar{g}_1,...,\bar{g}_n$) where $\bar{g}_i$ are the corresponding green times; and $\bar{q}_i$ are the flows which can be served during these fixed green times.

3 FEEDBACK-BASED GATING IN PRESENCE OF TIME-DELAY

The single gating location in Fig. 2(a) may be an ersatz for multiple gating locations, as in Fig. 3 later on, where only four out of the seven gated junctions are actually remote from the defined protected network. The gated vehicles at remote locations require a travel time to approach the PN. This travel time, imposes a time-delay (τ) in the action of the controller; therefore, this term should be considered in the design of the controller. How to obtain an ersatz model with one single gating location and one corresponding time-delay in cases of multiple real gating locations and corresponding different travel times, also designing a controller in the presence of time-delay are discussed following.

3.1 Model identification

In [16], a nonlinear continuous-time control design model of the protected network was developed based on physical considerations (using the travel time of gated flow, the conservation equation and the NFD). Eventually, the model was linearized around the nominal operation point (corresponding to $\hat{TTS}$), resulting in a first-order time-delay dynamic system as follows:

$$\Delta\dot{TTS}(t) = -A\Delta TTS(t) + B\Delta q_g(t-\tau) + B'\Delta q_d(t-\tau_d) \quad (4)$$

where Δ-quantities reflect corresponding deviations from nominal values; $T_c = 1/A$ is a time constant; B and B' are the model gains; and τ and τ_d are time-delays. Model (4) can be converted in discrete time (with sample time T) by use of standard formulas [24] as follows:

$$\Delta TTS(k+1) = \mu \cdot \Delta TTS(k) + \zeta \cdot \Delta q_g(k-m) + \gamma \cdot \Delta q_d(k-m_d) \quad (5)$$

where we have $A = -\ln(\mu)/T \sim (1-\mu)/T$; $B = \zeta/T$ and $B' = \gamma/T$; and $\tau \sim mT$, $\tau_d \sim m_d T$, with m , m_d integers.

According to well-known Control Engineering principles, the derived model structure in (5) is sufficient for selecting an appropriate feedback controller structure (in this case, a PI-type regulator, see Section 3.2), even without knowledge of the involved parameter values. The choice of the regulator parameters could then be effectuated manually, via trial and error. In contrast, if the parameter values of model (5) are known for a specific application, then the regulator parameters could be specified immediately, based on appropriately derived regulator design rules (Table 1). This Section 3.1 describes how to obtain the values of the model parameters in (5) for a specific application; while Section 3.2 develops and presents the design rules.

The derivation of an appropriate model for the control design purpose is called model identification. In Control Engineering, the field of System Identification uses statistical methods to build mathematical models of dynamic systems or to estimate the parameter values of available model structures from measured data (see, e.g. [25]). System Identification also includes the optimal design of experiments for efficiently generating informative data for fitting such models as well as model reduction and other relevant tasks.

In the present context, we need to develop an estimation procedure for the parameters of model (4) by use of measurements of the involved quantities, so as to enable a ready design of the gating feedback regulator. Measured values for $TTS(k)$ and $q_g(k)$ for a number of cycles k=1,…, K, may be assumed to be available, since they are needed for the operation of the feedback regulator. Note that, due to model linearization, the measured values should be collected so as to be in a neighborhood of the nominal conditions. Since the desired set-point $\hat{TTS}$ is also given, we may obtain corresponding values for $\Delta TTS(k) = TTS(k) - \hat{TTS}$, k=1,…, K; on the other hand, the exact corresponding nominal values for $\bar{q}_g$ and q_d are not known. A reasonable assumption is to calculate these values as averages, i.e.,

$$\bar{q}_{\mathrm{g}} = \frac{1}{K}\sum_{k=1}^{K} q_{\mathrm{g}}(k) \tag{6}$$

$$\bar{q}_{\mathrm{d}} = \frac{1}{K}\sum_{k=1}^{K} q_{\mathrm{d}}(k) \tag{7}$$

Using $\bar{q}_{\mathrm{g}}$ from eq. (6), we may derive values for $\Delta q_{\mathrm{g}}(k) = q_{\mathrm{g}}(k) - \bar{q}_{\mathrm{g}}$, k=1,…, K. However, there are no measurements available for the uncontrolled inflows q_{d}. Hence, as an approximation, we may use a constant term in place of $\gamma \cdot \Delta q_{\mathrm{d}}(k)$ in eq. (5); this constant term could be taken equal to the average of $\gamma \cdot \Delta q_{\mathrm{d}}(k)$; which, in view of eq. (7), turns out to be zero. In other words, we are assuming that the uncontrollable inflows reflected in $q_{\mathrm{d}}(k)$ do not exhibit strong variations near the nominal operation state; model inaccuracies resulting from this assumption act as disturbances to be rejected by the feedback regulator.

In conclusion, the parameter estimation procedure is carried out to specify appropriate values for μ and ζ for the system

$$\Delta TTS(k+1) = \mu \cdot \Delta TTS(k) + \zeta \cdot \Delta q_{\mathrm{in}}(k) \tag{8}$$

based on available data for $TTS(k)$ and $q_{\mathrm{in}}(k)$ as explained above. These data may be field data, in case of a field application, or simulation-based data, as in this paper. The pursued parameter estimation procedure is essentially a regression or curve fitting one, using least squares.

Generally, the method of least squares assumes that the best-fit curve of a given type is the curve that has the minimal sum of the deviations squared (least-square error) from a given set of data. Suppose that the data points are (x_1, y_1, z_1), (x_2, y_2, z_2), …, (x_K, y_K, z_K) where x and y are the independent variables and z is the dependent variable. In our case, the fitting curve is linear in the parameters and has the deviation (error) e from each data point, i.e., $e_1 = z_1 - (R \cdot x_1 + S \cdot y_1)$, $e_2 = z_2 - (R \cdot x_2 + S \cdot y_2)$, …, $e_K = z_K - (R \cdot x_K + S \cdot y_K)$. According to the method of least squares, the best fitting curve has the property that:

$$\Pi = e_1^2 + e_2^2 + \ldots + e_K^2 = \sum_{i=1}^{K} e_i^2 = \sum_{i=1}^{K} \left[z_i - (R \cdot x_i + S \cdot y_i) \right]^2 \rightarrow \underset{R,S}{\mathrm{Min}} \tag{9}$$

To derive R and S (i.e. the model parameters) we have to solve the system of equations

$$\frac{\partial \Pi}{\partial R} = 2\sum_{i=1}^{K} x_i \left[z_i - (R \cdot x_i + S \cdot y_i) \right] = 0 \tag{10}$$

$$\frac{\partial \Pi}{\partial S} = 2\sum_{i-1}^{K} y_i \left[z_i - (R \cdot x_i + S \cdot y_i) \right] = 0 \tag{11}$$

Which yields

$$R = \frac{\sum_{i=1}^{K} x_i y_i \cdot \sum_{i=1}^{K} y_i z_i - \sum_{i=1}^{K} y_i^2 \cdot \sum_{i=1}^{K} x_i z_i}{\left(\sum_{i=1}^{K} x_i y_i \right)^2 - \sum_{i=1}^{K} x^2 \cdot \sum_{i=1}^{K} y_i^2} \tag{12}$$

$$S = \frac{\sum_{i=1}^{K} x_i^2 \cdot \sum_{i=1}^{K} y_i z_i - \sum_{i=1}^{K} x_i y_i \cdot \sum_{i=1}^{K} x_i z_i}{\sum_{i=1}^{K} x_i^2 \cdot \sum_{i=1}^{K} y_i^2 - \left(\sum_{i=1}^{K} x_i y_i \right)^2} \tag{13}$$

In the case of our time-delay system (5), we have $z = \Delta TTS(k+1)$, $x = \Delta TTS(k)$ and $y = \Delta q_{\mathrm{g}}(k-m)$ where $m \triangleq$ is the time-delay. Since the least-squares procedure does not deliver best values for the time-delay m, the parameter estimation problem (for $R = \mu$ and $S = \zeta$) should be solved multiple times, for $m = 1, 2, \ldots$; the best obtained value for Π leads to the desired parameters. Consequently, μ, ζ and m may be derived and be used for the control design problem which is presented in the next section.

3.2 Controller design

The control objective in this study is to regulate the *TTS* (number of vehicles) in the PN, so as to maintain it around a selected set-point $\hat{TTS}$, via appropriate manipulation of the gated flow q_g in real time. The targeted set-point $\hat{TTS}$ should be selected within the range of critical values (i.e. yellow region in Fig. 1), in order to avoid throughput degradation. As mentioned earlier, [16] developed a control design model and an appropriate feedback controller for the described gating task. Given the derived model structure (5), the following proportional-integral-type (PI) feedback controller is appropriate

$$q_g(k) = q_g(k-1) - K_P\left[TTS(k) - TTS(k-1)\right] + K_I\left[\hat{TTS} - TTS(k)\right] \tag{14}$$

where q_g is the gated flow ordered by the controller, K_P and K_I are the proportional and integral gains, respectively. In case gating is applied at multiple junctions (as in this work), the flow ordered by (14) must be distributed among the gating positions (as illustrated in Fig. 2) via one, out of several possible, pre-specified policies (e.g. according to the respective saturation flows in this work). The regulator is constrained between maximum and minimum values. Beside this, upper and lower flow bounds are specified also for every individual gated link. If the regulator flow distribution is found to violate some of these individual bounds, then the surplus flows are re-distributed among the rest of the gated links. Note that gating is activated only within specific time windows (e.g. at the peak periods) or if some real-time measurement-based conditions are satisfied. Finally, the distributed flow to the gated links must be converted to appropriate green times, and the usual traffic signal settings in the corresponding junctions must be modified accordingly.

Generally, for good regulator gain values, either manual fine-tuning or some appropriate design rules may be utilized. It should be emphasized that feedback regulators are quite robust to moderate model parameter value changes. There are several practical design rules available to specify appropriate regulator gains for representative linear design models in a continuous-time setting. Such rules, however, seem to be lacking in the case of discrete-time systems; hence, we have developed a set of practical rules for the present general problem, which are summarized in Table 1. Specifically, for any real or simulated gating control problem, one needs first to gather measurements and estimate the model parameters μ, ζ and m, as detailed in Section 3.1. Based on these values, Table 1 delivers appropriate regulator gains K_P and K_I for the PI regulator (14). With this understanding, readers who are not familiar or not interested in the derivation of the entries of Table 1, may proceed directly to the next section.

The z-transform function of the process (5) is

$$P(z) = \frac{\zeta}{z^m(z-\mu)} \tag{15}$$

while the PI regulator (14) z-transform function is

$$C(z) = \frac{z\cdot(K_P + K_I) - K_P}{z-1} = K'\cdot\left(\frac{z - \frac{K_P}{K'}}{z-1}\right) \tag{16}$$

where $K' = K_P + K_I$.

By closing the loop with *C* (controller) from (16) and *P* (process transfer function) from (15) we have the following z-transform function for the closed-loop system:

$$F_C(z) = \frac{C(z)\cdot P(z)}{1 + C(z)\cdot P(z)} = \frac{K'\cdot\left(\frac{z - \frac{K_P}{K'}}{z-1}\right)\cdot\left(\frac{\zeta}{z^m\cdot(z-\mu)}\right)}{1 + K'\cdot\left(\frac{z - \frac{K_P}{K'}}{z-1}\right)\cdot\left(\frac{\zeta}{z^m\cdot(z-\mu)}\right)} \tag{17}$$

For $m = 0$, the corresponding values of Table 1 lead to $F_C(z) = 1/z$, i.e. to the fastest possible (dead-beat) control design, as also stated in [16]. Considering the time-delayed case ($m > 0$), a usual control

design step, which reduces the complexity (polynomial) of the closed-loop transfer function (F_C), is to specify the zero of the controller to be equal to the pole of the process, i.e. to set $K_P/K' = \mu$, which consumes a first degree of freedom in the control design. With this choice, the closed-loop transfer function (17) simplifies to

$$F_C = \frac{K'\zeta}{z^m(z-1)+K'\zeta} \tag{18}$$

The closed-loop transfer function (18) is identical with the one considered in Section 3.3 of [26], and hence, we can use the design rules of [26] to specify the second degree of freedom for our PI regulator for $m > 0$. Specifically, Figure 5 of [26] delivers the value of a parameter κ as a function of T/T_D. The correspondence between these quantities and our notation is $m = T_D/T$ and $\kappa = K' \cdot \zeta \cdot m$. Thus, for any $m > 0$, we can use Figure 5 of [26] to obtain the corresponding value of κ and hence of K'. This value, combined with our above choice $K_P/K' = \mu$, deliver the regulator parameter values displayed in Table 1.

For sufficiently long time-delays (i.e. $m \geq 4$), an alternative control design procedure for our discrete-time regulator is to consider a PI regulator design for the continuous-time system (4).Then, applying the well-known [26] rules to this system, we obtain very similar values for K_P and K_I as in Table 1. This confirms the pertinence of the previously derived rules, by use of a different design approach, for the cases with $m > 3$. Conducted simulations with different time delays m also confirmed the pertinence of the derived general rules of Table 1, which can be used as a black-box procedure in any related practical application.

4 TEST-BED DESCRIPTION

An extended part of the Chania urban road network, compared to ([16]; [17]), is modelled in in the AIMSUN microscopic simulation environment [28] according to Fig. 3 (blue dashed line demonstrates the extended part). The central business district (CBD) of the Chania urban road network, where the congestion usually starts during the peak period, is considered as the protected network (shown by the bold red line in Fig. 3) i.e. the urban network part to be protected from degradation due to over-saturation via appropriate gating actions at upstream links. Specifically, seven gating links are indicated by arrows in Fig. 3, three of them located at the border of the protected network (i.e. gated junctions 4, 6 and 7); and the four other located further upstream of the PN (i.e. junctions 1, 2, 3 and 5). In the middle of every link inside the red border line, a (emulated) loop detector has been installed, and the related measurements are collected at every cycle (in this case 90 s). To reduce the risk of obstructing the upstream junctions, the gating links are selected to provide sufficient space for queuing.

Multiple origins and destinations (O-D) are introduced at the network boundaries and at internal network locations, including the PN area, and are indicated by small circles in Fig. 3. The internal O-Ds mimic various corresponding in- and outflows, including on-street and off-street parking arrivals and departures, that may partially affect the PN area. When running the simulation with an O/D Matrix, a route choice system distributes the vehicles over the different paths from each origin to each destination. The distribution per path can depend on the cost of each link that composes the path. The cost of each link depends on the cost function associated to it. The default cost function in AIMSUN is travel time (selected also in this paper), but other cost functions for specific situations like tolls or, for

TABLE 1 Different values for K_P and K_I according to different time-delays (m)

m	K_P	K_I
0	μ/ζ	$(1-\mu)/\zeta$
1	$\mu/(3\zeta)$	$(1-\mu)/(3\zeta)$
2	$\mu/(5\zeta)$	$(1-\mu)/(5\zeta)$
3	$\mu/(6\zeta)$	$(1-\mu)/(6\zeta)$
>3	$\mu/(2m\zeta)$	$(1-\mu)/(2m\zeta)$

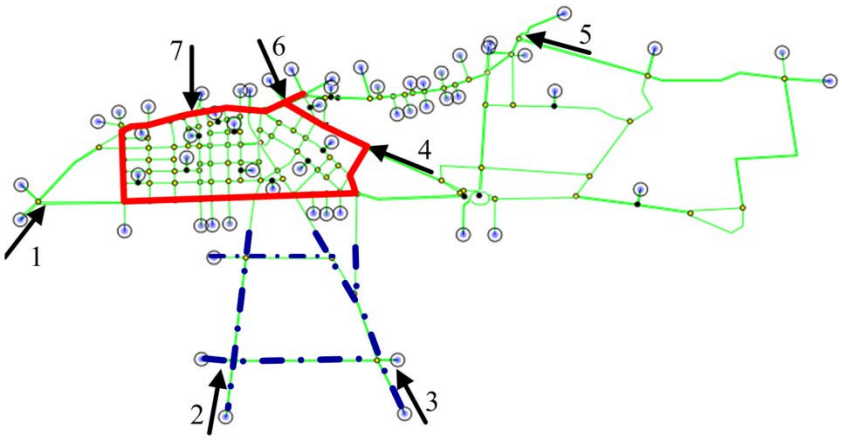


FIGURE 3 Chania urban network modelled in AIMSUN, blue dashed line extended compared to [16]

instance, cost functions that take into account other factors like the psychological weight of low speeds can be selected. A time step of 30s is chosen as the route choice period in this study. In particular, if gating measures create long queues and delays at the gated links, alternative routes (if available) may be selected by the drivers towards their respective destinations; clearly, this reflects the medium-term routing behaviour of drivers to any introduced gating measures. Note also that this diversion may jeopardize to some extent the intended gating impact if drivers divert and enter the PN via non-gated links; therefore, the choice of gating links should also consider the availability and potential attractiveness of alternative routes that bypass the gating location.

The real peak-hour fixed-time signal plan of the Chania urban road network is applied in the simulations without gating. This plan is modified at the gating junctions when real-time gating actions are present, by modifying accordingly the involved signal stage durations. Specifically, the gating strategy determines the duration of one (gated) signal stage (within pre-specified bounds) in real-time; any (positive or negative) deviation of this stage duration from its fixed-time value is assigned to other stages of the same junction.

5 SIMULATION RESULTS

Due to the stochasticity of the microscopic simulator AIMSUN, different simulation runs (replications) with different random seeds may lead to different results. Considering the stochastic and chaotic behavior in real traffic networks, small changes in initial or intermediary conditions may lead also to different outcomes. Here, 10 replications for each investigated scenario have been carried out and then the average value of the 10 runs for each performance index is calculated. Three performance indexes are used here (as provided by AIMSUN): the average vehicle delay per km, the mean speed and the total number of vehicles that exit the network. Note that these indexes are provided for the entire Chania urban network (not only the PN), thus the delays at the gated links are actually included in the performance indexes.

5.1 NFD of simulated PN

Fig. 4(a) and (b) display the complete operational NFD (90 s cycle-based measurement points) obtained for the PN of Fig. 3 for the loading (first 2 hours) and the whole (i.e. 4-hour) AIMSUN simulation period (including the network recovery period, indicated by +), respectively. As pointed out also in Section 4, the DTA based routing is activated during the simulation runs; ten different replications (each with different seed in AIMSUN) of the 4-hour scenario were carried out and are included in Fig. 4. A trapezoidal traffic demand profile, starting from very low values and increasing gradually to levels that result in congestion within PN (as under typical real traffic conditions at the peak periods) is introduced to the network; ultimately, the demand is gradually reduced, until the network is virtually emptied at the end of the simulation (see Fig. 4(b)). Fig. 4(a) demonstrates that a

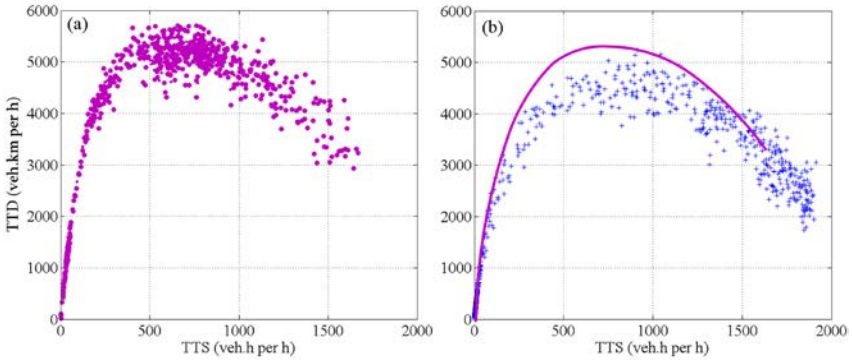


FIGURE 4(a) NFD for loading PN for 10 replications; (b) **NFD for loading and unloading for 10 replications (the pink bold line is the fitted curve to the scatterplot in (a)**

fundamental diagram (asymmetric inverse-U shape) is indeed occurring during the 2-h network filling period, with low scatter even across different replications; Fig. 4(b) indicates that the inverse-U shape appears also during the decreasing demand period of 2 h, albeit with a visible hysteresis compared to the filling 2-h period. The maximum *TTD* values in the diagram occur in a *TTS* region of 500 to 800 veh·h per h. These phenomena may be exploited for gating control by maintaining the *TTS* in the aforementioned range, so as to maximize the throughput in the PN. An uncontrolled *TTS* may lead to the decrease of *TTD* (and hence of the PN throughput), if the optimal range of *TTS* is exceeded.

5.2 Non-gating case

As pointed out before, the traffic signal plan in the non-gating case is set according to the fixed-time signal settings utilized in the real Chania network during the peak period. Table 2(a) displays the aforementioned indexes for every replication (R). The last two columns display the average and standard deviation (S. D.) for each index. The standard deviation is considered as an index of reliability of the traffic conditions over different runs of simulation. It is observed that the non-gating scenario leads to high S.D. for both delay and mean speed, i.e. 84 (s/km) and 2.2 (km/h), respectively.
The detailed results of one specific replication (R10) are displayed in Fig. 5, for an illustrative comparison of the non-gating versus gating cases. Fig. 5(a) and (d) display the PN's *TTS* and the total flow served by the gated junctions (from the 7 gated links) q_g and the PN's *TTD* are shown in (Fig. 5(b) and (h)) and (Fig. 5 (c) and (i)), respectively. Note that q_{in} and q_g are different in the present application, as gating is applied partly further upstream (gating points 1, 2, 3 and 5) of the PN perimeter. Focusing on the left column of Fig. 5 (i.e. (a), (b), (c)), with increasing demand, all three displayed quantities are gradually increasing, as typical in under-saturated conditions. At time $t = 0.6$ h, the increased demand leads to the observed surge of *TTS* and *TTD*, the latter reaching soon capacity values according to Fig. 4 (a), while the former is traversing the aforementioned critical region of [500, 800]. As the demand keeps increasing, *TTS* continues to increase and reaches very high values (2000 veh·h per h); as a consequence, link over-spilling and gridlock phenomena appear in the PN and lead to a sudden drop in *TTD* values (see Fig. 5 (c)), that are persisting until about $t = 2.6$ h; eventually, in the period $t \in [2.6\text{h}, 3.3\text{h}]$, *TTD* increases again thanks to the reduced congestion due to the decreasing demand at the end-stage of simulation.

5.3 Gating (with time-delay)

In this section, the simulation results while applying the gating control strategy to the same Chania network as for the non-gating case are presented. The set-point for the regulator may be taken from the NFD displayed on Fig. 4 . Specifically, the utilized $\hat{TTS}$ value is 750 veh·h per h. The utilized regulator parameters K_P and K_I were derived as in section 5.3.1.

5.3.1 Deriving the gain values (K_P and K_I)

The utilized regulator parameters K_P and K_I were obtained after a least square estimation by applying the time-series of (q_g, *TTS*) data around the *TTS*-critical (750 veh·h per h) and considering $m = 3$. To specify m, the longest path from the remote gating positions (i.e. junctions 1, 2, 3 and 5 in Fig. 3) to the PN is found. The distance from junction 5 to the PN border is about 900 m which, by considering an average speed of 12 km/h (including waiting times at three signalized junctions in-between and the congestion occurring in this path during the peak period) for the vehicles traveling to the PN, leads to a value of 270 s as the travel time of a vehicle to reach the PN. Consequently, the gating action may face a delay of up to 270 s, or three signal cycles, to reach the PN. This time-delay of *m*=3 was considered first in the least-squares procedure. Thus, in the first step, the model (5) parameters were estimated to $\mu = 0.769$ and $\zeta = 0.012$; and by applying Table 1 for $m = 3$ we derive $K_P = 10\,\text{h}^{-1}$ and $K_I = 3\,\text{h}^{-1}$. In cases where the gating positions are located at different distances from the PN (as in this study), the aforementioned procedure might turn out to be conservative, because, the longest travel time to the PN is considered as the time-delay of the system. Moreover, the farthest gating junction may not be the most influential among the other gating positions, since this depends highly on the demand profile and the flow which is served in that gated junction. Thus, one may use the full method discussed in Section 3.1, and carry out the estimation procedure for different values of *m*; then we obtain for $m = 1$: $\mu = 0.812$, $\zeta = 0.023$ and $\Pi = 29429$; for $m = 2$: $\mu = 0.781$, $\zeta = 0.027$ and $\Pi = 33704$; for $m = 3$: $\Pi = 23674$ and the aforementioned values for μ and ζ. Consequently, the minimum value for Π is derived indeed from the parameter estimation with $m = 3$.

5.3.2 Time-delayed gating results

Fig. 5 (second column) displays the results for the gated case (same replication as for non-gating), which can be directly compared to the non-gating results displayed in the first column of the same figure. Up to around $t = 1.1$ h, the traffic conditions are identical as in the non-gating case. When gating is activated (shown by the dashed violet vertical line in Fig. 5(h)), as *TTS* approaches its set-point ($\hat{TTS} = 750$), the gating regulator orders lower gating flow values to maintain *TTS* around its set point, and, as a consequence, *TTD* is maintained at high levels (see Fig. 5(i)), avoiding the discussed drop of *TTD* in time period $t \in [1.8\text{h}, 2.6\text{h}]$ in the non-gating scenario (see Fig. 5(c)). It is visible in Fig. 5(h) that the served gated flow, shown by the red line, deviates at times from the values ordered by the regulator (shown by the black line). A thorough inspection of the total served gated flow (not shown here) revealed that this gap is mainly due to some specific gated links, whose demands are not always sufficient to create the respective inflows assigned to them. As it was expected, thanks to the chosen feedback structure of the controller, this deviation has virtually no influence on the regulator's performance for keeping PN's *TTS* close to the pre-specified set-point. As we reach the end period of the simulation, *TTS* is returning to lower values, and the relative difference between the actual and ordered gated flows grows higher, hence gating is switched off, and the traffic flow returns to under-saturated conditions due to lower demand. Interestingly, the congestion period in the gated scenario is reduced by almost 40 minutes (see Fig. 5(a) and (g)) compared to the non-gating case.

TABLE 2 (a) performance indexes results for non-gated scenario; (b) results for the gated scenario with time-delay (*: in 1000); (c) results for the gated scenario with bigger control steps (Non-time-delayed)

Scenario	Indexes	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	Ave.	S. D
(a) Non- gated	Delay (s/km)	230	260	285	421	405	352	187	413	214	369	314	84
	Speed (km/h)	11.6	10.6	9.9	7.2	7.4	8.3	13.4	7.3	12.2	8.0	9.6	2.2
	Vehicles out*	12.8	12.9	12.6	12.8	12.8	13.0	12.8	12.8	12.7	12.8	12.8	-
(b) Gated with time-Delay	Delay (s/km)	213	232	189	214	202	220	180	226	187	206	206.9	16.5
	Change (%)	-7.4	-10.8	-33.7	-49.2	-50.1	-37.5	-3.7	-45.3	-12.6	-44.2	**-34.0**	**-80.4**
	Speed (km/h)	12.2	11.4	13.3	12.1	12.7	11.9	13.8	11.7	13.4	12.5	12.5	0.8
	Change (%)	5.2	7.5	34.3	68.1	71.6	43.4	3.0	60.3	9.8	56.3	**30.3**	**-65.1**
	Vehicles out	12.8	12.9	12.6	12.8	12.8	13.0	12.8	12.8	12.8	12.8	12.8	-
(c) Gated without time-delay	Delay (s/km)	184	189	192	181	185	197	187	219	188	212	193.4	11.4
	Change (%)	-20.0	-27.3	-32.6	-57.0	-54.3	-44.0	0.0	-47.0	-12.1	-42.5	**-38.3**	**-85.9**
	Speed (km/h)	13.6	13.2	13.2	13.7	13.5	12.9	13.4	12	13.4	12.3	13.1	0.5
	Change (%)	17.2	24.5	33.3	90.3	82.4	55.4	0.0	64.4	9.8	53.8	**36.8**	**-75.2**
	Vehicles out*	12.8	12.9	12.6	12.8	12.8	13.0	12.8	12.8	12.8	12.8	12.81	-

Table 2(b) displays the achieved gating improvements of the average performance indexes over the non-gating case of Table 2(a). As expected, because of the identical traffic demand in both scenarios and the fact that the network is evacuated at the end of the simulation, the total number of vehicles served in both cases is close to each other. There are remarkable improvements in the overall mean speed and the average delay in order of 30% and 34%, respectively. Table 2(b) reports also huge reductions of S. D., i.e. 80% for the delay and of 65% for the speed, compared to the non-gating case.

5.4 Gating (without time-delay) by increasing control step

Further investigation on the gating control action, which was not conducted in past works, was has been carried out by increasing the control step for the regulator (14). For the same scenario discussed in the previous section, the gating control action is implemented with a time step of five cycles (i.e. 450 s) in place of one cycle (i.e. 90 s). In other words, the gating action is updated with new measurements every five cycles. By considering this time step, each time the regulator orders a flow, this flow is implemented for five consecutive cycles, before it is updated again. Given this long time step, we may assume that the control system is operating with virtually zero time-delay.

5.4.1 Deriving the gain values (K_P and K_I)

By using time-series of (q_g, *TTS*) data (now averaged over 5-cycle periods) around the *TTS*-critical (750 veh here) and considering *m*=0 for the estimation procedure, we end up with $\mu = 0.760$ and $\zeta = 0.011$. By virtue of Table 1, for *m*=0, we have $K_P = \mu/\zeta = 65$ and $K_I = (1-\mu)/\zeta = 20$.

5.4.2 Non-time-delayed gating results

Table 2(c) displays the detailed simulation results for the gated scenario with longer time step. Interestingly, the results indicate similar efficiency compared to the simulation results in the previous section. It should be emphasized that this is an important finding, especially for the case of real-filed implementation. This study proves that for an efficient gating control, the control action can be executed even in a bigger time step. According to Table 2(c), significant improvements in the overall mean speed and the average delay in order of 36.8% and 38.6%, respectively, have been achieved.

The results of the non-gated (first column: (d)-(f)) and gated scenario (second column: (j)-(l)) are demonstrated in Fig. 5 for replication 10. In fact, the non-gated scenario in the first column is the same replication as the one reported in section 5.2 (but with lower time resolution, 450s) and was discussed in details in that section. The second column displays the results for the gated case (same replication as for non-gating), by considering the aforementioned parameters in the regulator's equation. As expected, the regulator attempts to maintain the *TTS* of the PN (see Fig. 5(j)), close to the pre-specified set-value (i.e. 750 veh.). After activation of gating control, (shown by the dashed violet vertical line in Fig. 5(j)), as *TTS* approaches its set-point, the gating regulator orders lower gating flow values (2200 veh./h) to keep *TTS* around its set point, and, as a consequence, *TTD* is maintained at high levels (see Fig. 5(l)). Since the control step is five cycles in this case, the activation of the gating strategy (i.e. exceeding 85% of $\hat{TTS} = 750$ after 2 cycles, defined in the algorithm), starts slightly later than in Fig. 5 (h). However, the controller acts perfectly and keeps the *TTS* pretty close to the critical value. After applying the gating control, the overall delay and mean speed are reduced to 212 s/km and 12.3 km/h, respectively.

This study proved that the real-time gating action can be executed at a slower pace. In other words, this finding shows that even with a bigger control step (not one signal cycle as in all previous studies), the gating strategy acts efficiently and leads to significant improvements in mean speed and delay at the overall network level. This may be very interesting for the urban traffic management authorities; because it shows that the signal plan can be modified less frequently during the gating action.

6 CONCLUSION

This study demonstrated that efficient feedback-based gating control may be applied not only at the boundary, but also further upstream of the PN. Based on restrictions which might be faced for implementing gating control at the perimeter of the PN in different urban networks, such as unavailability of proper links to store the gated vehicles (queuing) or sufficient number of signalized junctions, the traffic may be also metered at some junctions remote from the border of the PN. For the

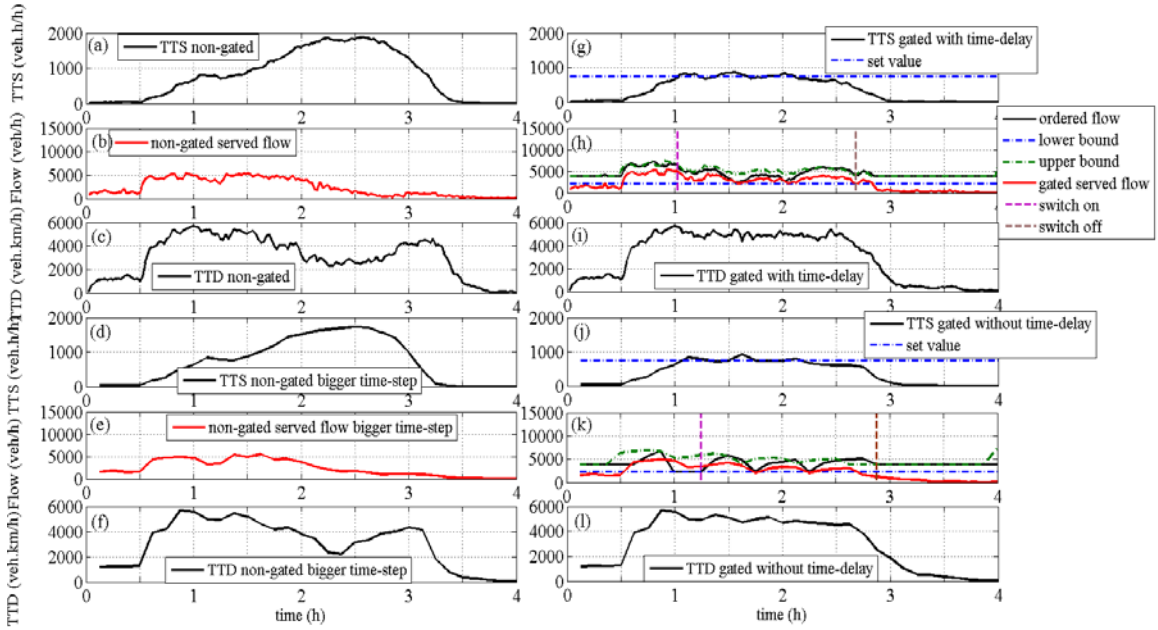


FIGURE 5 (a) PN's *TTS* vs. time in non-gating case; (b) total served flow by the gated junctions vs. time for the non-gating case; (c) PN's *TTD* vs. time for the non-gating case; (d) PN's *TTS* vs. time in non-gating case with bigger control step; (e) total served flow by the gated junctions vs. time for the non-gating case; (f) PN's *TTD* vs. time for the non-gated case with bigger time-step; (g) PN's *TTS* vs. time for the gating case with time-delay (h) served and ordered flow vs. time for the gating case; (i) PN's *TTD* vs. time for the gated case with time-delay; (j) PN's *TTS* vs. time for the gating case; (k) served and ordered flow vs. time for the gating case with bigger control step; (l) PN's *TTD* vs. time with the gated case for bigger control step

gating procedure in this study, the recently proposed feedback gating concept is employed by exploiting the operational NFD of the PN.

General procedures are proposed for deriving a control design model and estimate its parameters based on measured data; and for selecting proper regulator parameters, without the need for manual testing and fine-tuning. This enables the easy application of the feedback gating concept for a large variety of problem settings and infrastructures. The concept is demonstrated in a realistic simulation environment and significant improvements in terms of mean speed and average delay are achieved by applying the gating strategy.

This paper also demonstrates that the real-time gating action can be executed at a slower pace. In other words, this finding shows that even with a bigger control step (bigger than one signal cycle as in all previous studies), the gating strategy acts efficiently and leads to significant improvements in mean speed and delay at the overall network level. This may be very interesting for the urban traffic management authorities; because it shows that the signal plan can be modified less frequently during the gating action.

Further research directions include comparison with more comprehensive traffic-responsive signal control strategies, designing a hybrid control strategy which combines the gating concept with a traffic responsive control strategy downstream of the gated links (to avoid spillback) and field implementation of the concept.

ACKNOWLEDGMENT

The research leading to these results has been partly funded by the European Commission FP7 program NEARCTIS (Network of Excellence for Advanced Road Cooperative Traffic management in the Information Society) and the Netherlands organization for scientific research (NWO) in the grant "there is plenty of room in the other lane".

REFERENCES

[1] Hunt, P.B., Robertson, D.I, Bretherton, R.D., Royle, M.C., 1982. The SCOOT on-line traffic signal optimization technique. Traffic Engineering and Control 23, 190-192.
[2] Lowrie, P.R., 1982. SCATS: the sydney co-ordinated adaptive traffic system-priciples, methodology, algorithms. IEEE International Conference on Road Traffic Signalling (67-70), London, England.
[3] Lo, H.K., Chang, E., Chan, Y.C., 2001. Dynamic network traffic control. Transportation Research Part A 35, 721–744.
[4] Chang, T.H., Sun, G.Y., 2004. Modeling and optimization of an oversaturated signalized network. Transportation Research Part B, 687-707.
[5] Wood, K., Bretherton, D., Maxwell, A., Smith, K., Bowen, G., 2002. Improved traffic management and bus priority with SCOOT (TRL Staff Paper PA 3860/02). London, UK: Transport Research Laboratory.
[6] Godfrey, J. W., 1969. The mechanism of a road network. Traffic Engineering. Control 11, 323–327.
[7] Geroliminis, N., Daganzo, C.F., 2008. Existance of urban-scale macroscopic fundamental diagrams: Some experimental findings. Transportation Research Part B 42, 756-770.
[8] Buisson, C., Ladier, C., 2009. Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams. Transportation Research Record 2124, 127–136.
[9] Gayah, V.V., Daganzo, C.F., 2011. Clockwise hysteresis loops in the macroscopic fundamental diagram: An effect of network instability. Transportation Research Part B 45, 643-655.
[10] Daganzo, C. F., Gayah, V. V., Gonzales, E. J. (2011). Macroscopic relations of urban traffic variables: Bifurcations, multivaluedness and instability.Transportation Research Part B: Methodological, 45, 278-288.
[11] Knoop, V.L., Hoogendoorn, S.P., Van Lint, J.W.C., 2013. Impact of traffic dynamics on macroscopic fundamental diagram. Presented at the 92nd Annual Meeting of the Transportation Research Board.
[12] Daganzo C. F., 2007. Urban gridlock: Macroscopic modeling and mitigation approaches. Transportation Research Part B 41, 49-62.
[13] Zhang, Y., Bai, Y., Yang, X.G., 2010. Strategy of traffic gridlock control for urban road network. China Journal of Highway Transportation, 23, 96-102.
[14] Geroliminis, N., Haddad, J., Ramezani, M., 2012. Optimal perimeter control for two urban regions with macroscopic fundamental diagrams: A model predictive approach. IEEE Transaction on Intelligent Transportation Systems, vol. 14, 348-359.
[15] Lia, Y., Xua, J., Shena, L., 2012. A Perimeter Control Strategy for Oversaturated Network Preventing Queue Spillback. Procedia-Social and Behavioral Sciences 43, 418-427.
[16] Keyvan-Ekbatani, M., Kouvelas, A., Papamichail, I., Papageorgiou, M., 2012. Exploiting the fundamnetal diagram of urban networks for feedback-based gating. Transportation Research Part B 46 (10), 1393–1403.
[17] Knoop, V.L., Van Lint, J.W.C., Hoogendoorn, S.P., 2012. Routing strategies based on the macroscopic fundamntal diagram. Transportation Research Records 2315, 1-10.
[18] Keyvan-Ekbatani, M., Papageorgiou, M., Papamichail, I., 2013a. Urban congestion gating control based on reduced operational network fundamental diagrams . Transportation Research Part C 33, 74-87.
[19] Keyvan-Ekbatani, M., Yildirimoglu, M., Geroliminis, N., Papageorgiou, M., 2013b. Traffic Signal Perimeter Control with Multiple Boundaries for Large Urban Networks, Proc. 16th IEEE Conference on Intelligent Transportation Systems (ITSC 2013), The Hague, The Netherlands.
[20] Aboudolas, K., Geroliminis, N., 2013. Perimeter and boundary flow control in multi-region heterogeneous networks. Transportation Research Part B 55, 265–281.
[21] Geroliminis, N., Zheng, N., Ampountolas, K., 2014. A three-dimensional macroscopic fundamental diagram for mixed bi-modal urban networks. Transportation Research Part C 42, 168-181.
[22] Aboudolas, K., Zheng, N., Geroliminis, N., 2014. Perimeter flow control for bi-modal urban road networks. 93th TRB Annual Meeting, Washington, DC, USA.
[23] Vigos, G., Papageorgiou, M., 2010. A simplified estimation scheme for the number of vehicles in signalized links. IEEE Trans. Intelligent Transportation Systems 11, 312-321.
[24] Seborg, D., Edgar, T.F., Mellichamp D.A., 1989. Process Dynamics and Control. New York: Wiley.
[25] Ljung, L., 1987. System Identification: Theory for the User. Prentice Hall.
[26] Papageorgiou, M., Messmer, A., 1985. Continious-time and discrete-time design of water flow and water level regulators. Automatica 21(6), 649-661.
[27] Chien, K. L., Hrones, J. A., Reswick, J. B., 1952. On the Automatic Control of Generalized Passive Systems. Transactions of the American Society of Mechanical Engineers 74, 175-185.
[28] TSS transport simulation systems, 2008. AIMSUN User Manual Version 6, Barcelona, Spain.