Abstract—Many network-wide traffic management strategies based on the Macroscopic or Network Fundamental Diagram (NFD) have been developed in recent years. In this paper we investigate one of these strategies, namely feedback NFD-based perimeter control, which limits the rate at which vehicles are allowed to enter an urban region. The inflow limitation is imposed at selected main entry links (gated links) and results in queuing of vehicles at the boundaries of the network. Most of the works on this subject neglected the effect of the queued vehicles on the traffic conditions upstream of the gated links. In this paper we analyze in microsimulation the queuing at the gated links and its effect on network performance for a realistic network model. Moreover, a queue management strategy based on a continuous quadratic knapsack problem aiming at balancing relative queues at the gated links is proposed. Simulation results suggest that queues under perimeter control are shorter in space and time than with no perimeter control. Additionally, managing the queues at the gated links not only improves the overall network performance but also reduces the possibility of queue propagation to the upstream junctions. This improves traffic flow outside the protected network.

I. INTRODUCTION

Extensive literature exists on modeling vehicular traffic dynamics on large urban networks (for a review see [1]). The advent of the concept of the Macroscopic or Network Fundamental Diagram (MFD or NFD) and the demonstration of its existence with field data provided new opportunities for the modeling of urban networks [2]. As a consequence, a series of new NFD-based urban traffic management strategies emerged: perimeter urban traffic flow control [3], [4], [5], some considering the presence of public transport [6] or freeways [7]; traffic routing [8], [9]; parking pricing [10]; and congestion pricing [11].

Among the proposed NFD-based strategies, perimeter urban traffic control is particularly attractive because it may benefit from existing traffic control system infrastructure such as traffic loop detectors, traffic controllers, and traffic signals. Perimeter control attempts to sustain traffic conditions that maximize network throughput by holding vehicles outside of a protected network (PN) area via reduced green times at main entrances, the so called gated links.

A veiled criticism about NFD-based studies on perimeter traffic control refers to the fact that simulated networks were of limited size and did not model an extensive area around the PN; or did not consider cases of interference of the created queues at the gated links with upstream junctions and the respective impact on delays which may lead to overestimating the benefits of the approach. A notable exception is the work in [12] that integrates the boundary queue dynamics to an MFD-based model. The model is then applied for perimeter optimal control on a two-region MFD system so as to avoid queue interference with the MFD dynamics, i.e., preserve well-defined MFDs on both regions.

In this paper we present preliminary results of the analysis of the resulting queues and delays at the gated links in the simulation of a realistic network subject to NFD-based feedback perimeter control. We show that the created queues without perimeter control are longer in space and time than with perimeter control. This comes with no surprise, since perimeter control keeps the network operating around its capacity. Indeed, it is well known in freeway traffic flow control that temporary ramp metering queues [13] or mainstream traffic flow control queues [14] are the (low) price to be paid for higher throughput compared to the (congested) no-control case; and that traffic control may eventually reduce, in average, those queues as well. Furthermore, we propose a strategy for the balancing of relative queues at the gated links that may reduce the interference with upstream junctions and the resulting delay when perimeter control is under operation.

In Section II, we review the basics of NFD-based feedback perimeter control. Section III presents the queue management strategy proposed in this paper. Queues and delays at the gated links, as well as network performance and the queue management strategy, are evaluated in Section IV. Finally, Section V concludes the paper.

II. NFD-BASED FEEDBACK PERIMETER CONTROL

Traffic management via perimeter control may be realized by the use of a feedback regulator designed after a control design model based on the concept of the NFD. In this section, we introduce some definitions related to the NFD and present the feedback regulator that delivers the total flow that should enter the PN. A discussion on how the total flow may be distributed among the gated links is also presented.
A. Network Fundamental Diagram (NFD)

The NFD describes a relation between the network production (or total weighted flow) and accumulation (or average density) under homogeneous traffic conditions [2]. Evaluating the NFD of a given network allows one to determine the maximum number of vehicles that should be maintained inside that network so as to have maximum throughput, i.e., operate at the diagram’s critical region.

In this paper we follow [3] and represent the NFD alternatively as the relation between Total Traveled Distance, $TTD$ (veh·km/h), and the Total Time Spent, $TTS$ (veh·h/h). This form is suitable for use with the available traffic measurements. The $TTD$ and $TTS$ for a given network are derived via the following equations [3]:

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

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

with $k = 0, 1, 2, \ldots$ the discrete time-step reflecting the corresponding traffic cycles with length $T$ (h), $M$ the set of network links equipped with detectors, and $q_z$ (veh/h) and $L_z$ (km), the measured flow and length of link $z$, respectively. Note that in this case the $TTS$ corresponds to the total number of vehicles $\hat{N}$ estimated from all equipped links. The estimated number of vehicles in link $z$ is given by

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

with $o_z$ (%) and $\mu_z$ (lanes), the occupancy measurement and the number of lanes of link $z$, respectively, and $\lambda$ (km) the average vehicle length.

B. Feedback Perimeter Traffic Flow Control

A feedback Proportional-Integral (PI) regulator may be used to determine the amount of flow $q_g$ that should enter the network through the gated links during the next time period [3], and is given by

$$q_g(k) = q_g(k - 1) - K_P [TTS(k) - TTS(k - 1)] + K_I \left[ TTS - TTS(k) \right]$$ (4)

with $K_P$ and $K_I$ the proportional and integral gains, respectively, and $TTS$ a set-point value within the NFD critical region for network throughput maximization. The $TTS$ value used as the input of the regulator at every time step $k$ is estimated from occupancy measurements via (3) and (1). More details about the design and implementation of (4) may be found in [3, 5].

C. Flow Distribution

The flow $q_g$ resulting from (4) corresponds to the total flow that should enter the PN through the $n$ gated links and must be appropriately distributed among them.

A possible approach is to distribute the flow proportionally to each gated link according to its saturation flow as, e.g., in [3]. This policy may result in wasted green times during activation of the perimeter control. However, its integration with a spill-back and low demand (at the gated links) alert algorithm reduced the difference between the ordered and actual (or served) flow and consequently lowered the possibility of wasting green times [5].

Other possibilities for distributing the total flow may aim at balancing the relative link queues, or delays or waiting times at the gated links; as also done in [15] for a dual branch freeway on-ramp, in [16] for merging freeways, and in [17] for integrated freeway traffic flow control.

Whatever distribution method is pursued, the flows $q_i$ to be implemented at each gated link $i$ must satisfy

$$\sum_{i=1}^{n} q_i = q_g$$ (5)

for a successful perimeter control operation.

Because traffic lights are employed to regulate flow at the gated links, the distributed flows must be translated to green times based on the links’ characteristics such as lost time and saturation flow. Moreover, typical restrictions on the signal settings must be respected, thus the flow on a gated link may be subject to a lower bound $q_{\min,i}$ (veh/h), e.g., due to a specified minimum green time, and to an upper bound $q_{\max,i}$ (veh/h), e.g., due to the green time duration of the respective phase of the base signal plan, i.e.,

$$q_{\min,i} \leq q_i \leq q_{\max,i}.$$ (6)

The flow distribution, as we shall see later, should be designed carefully since it has a direct effect on the links’ queue length and waiting times. A limited storage capacity of a link may result in blocking of upstream junctions and a corresponding detrimental effect on system delay. Additionally, the flow distribution may influence the homogenization of network flows, thus affecting the NFD and, as consequence, the quality of the control operation.

III. QUEUE MANAGEMENT

In this section we develop a queue management strategy for the gated links. Queues formed at gated links might exceed the link length and interfere with upstream junctions at some locations, while the link storage capacity (in number of vehicles) is underutilized at other locations. From that perspective, queue lengths could be taken into account as a mechanism for deciding on flow distribution (see Section II-C) so as to make a better use of links storage space, thus reducing the effect on the traffic conditions at the surrounding areas and improving even more the overall network performance.

A. The Queue Model

Assuming there are no sinks or sources within the gated link, the time evolution of the queues $N_i$ (veh) on each gated link $i$, $i = 1, 2, \ldots, n$, may be derived from the equation of the conservation of the number of vehicles [15] as

$$N_i(k + 1) = N_i(k) + T[d_i(k) - q_i(k)]$$ (7)
in which $d_i$ (veh/h) is the flow that enters and $q_i$ (veh/h) is the flow that leaves the gated link $i$ during time period $[kT, (k+1)T]$.

We can rewrite (7) as

$$N_i(k+1) = A_i(k) - B_i(k)q_i(k)$$

with

$$A_i(k) = N_i(k) + Td_i(k)$$
$$B_i(k) = T.$$

### B. Relative Queue Balancing

The ordered flow $q_i$ may be distributed among all gated links aiming at balancing relative link queues. A relative queue value is obtained by dividing the current queue length $N_i(k)$ by the maximum queue length $N_{\text{max},i}$ (veh) defined for the respective gated link $i$. With balanced relative queue lengths all queues will reach their respective maxima virtually emptied by the end of the simulation.

### A. Network Description

Figure 1 (left) depicts the simulated network that corresponds to a portion of the city of Chania, Greece. The area within the dashed box is enlarged on the right side of Fig. 1 and corresponds to the city center of Chania. The shaded area corresponds to the PN. Note that, compared to the real network an extra junction has been added upstream of the gated link 2, to assess the impact of queue propagation to the upstream links.

The PN consists of some 80 junctions, 27 of which are signalized with traffic lights, and 165 links. Many of those links may experience over-saturated conditions in the peak-period. The traffic signals operate on the basis of fixed time periods we say that perimeter control is activated. In this case, the green splits of the fixed time plans at the junctions with gated links are overridden by the regulator, otherwise the fixed-time signal plans remain unchanged.

For the implementation of the queue management and performance analysis, the average queue lengths ($N_i$) and delays at the gated links were collected directly from AIMSUN. In a real setting these may be estimated from loop detector measurements by various techniques, see [15] and references therein. The inflows ($d_i$) at the entrance of the gated links were collected from loop detectors. The maximum storage space of each gated link $i$, in number of vehicles, has been considered as the maximum queue length value ($N_{\text{max},i}$).

The dynamic traffic assignment option in AIMSUN has been activated (see [3]) in order to obtain a more realistic and homogeneous distribution of demand within the network.

### D. Analysis

In this paper, delay, expressed in s/km, is chosen as the metric of network performance. It indicates the increase of travel time compared to free flow conditions where traffic is uninterrupted by traffic lights. The total delay results from the multiplication of the delay with the total traveled distance. All results are based on traffic variables measured during a sample time of 90 s.

### E. Results

For each scenario, ten replications were carried out in order to account for the stochastic effects of the microscopic simulation. The box plot displayed in Fig. 2 summarizes the overall network delay values (all replications) for the three scenarios (NPC, PC and PCQ) in terms of minimum and
maximum (top and bottom black bars), first quartile and the third quartile (top and bottom of the blue box), and median (red bar inside the box). As it can be observed in the figure, the delay is remarkably reduced in the PC and PCQ scenarios (in average 34% and 39%, respectively). In addition, a smaller height of the box (representing the interquartile range (IQR)) can be seen in PC and PCQ scenarios which indicates lower variability in the results of these scenarios compared to the NPC scenario. In the NPC scenario, the spread is rather large which is typical for heavily congested traffic, whereby small stochastic variations may lead to excessive queues in some parts of the network spilling back to the first upstream junction and interfering with the traffic flow there.

1) Analysis of the NFDs: The NFDs of the PN for the three scenarios (all replications) are presented in Fig. 3. For a better distinction, the measurement points are displayed with black dots for the network loading (first 2 hours) and with gray dots for the recovery period (last 2 hours).

The maximum network throughput (or TTD) of around 5400 veh·km/h can be observed to occur at a TTS range of 500–750 veh·h/h in the NFD of the NPC scenario in Fig. 3(a). Recall that the value of $TTS$ was chosen within this range for maximum throughput (see Section IV-C).

In Fig. 3(a) a smoothed line connecting the points of the NFD for a single replication is shown. The low TTD values attained prevent the network from recovering before the end of the simulation. This is due to local gridlocks that keep part of the network congested whereas other parts free gradually, leading to a lower production rate (incomplete) recovery.

Figure 3(b) shows the NFD for the PC scenario. The regulator succeeded to maintain the TTS close to $TTS$ during the perimeter control activation, showing that the approach is effective. Therefore, perimeter control contributes to keeping the production at a high level and, hence, the delays low (see Fig. 2). A typical hysteresis loop (for loading and unloading the network) is visible more clearly in this scenario [20].

Finally, the NFD for the PCQ scenario is displayed in Fig. 3(c). As in the PC scenario (Fig. 3(b)), the regulator tries to keep the throughput at the maximum level. Unlike the NFD for PC, however, the critical accumulation ($TTS = 600$ veh·h/h) is exceeded in some replications. This is the result of the attempt of queue management to balance relative queues at the gated links. The resulting flow distribution may lead to higher inflow into the PN at junctions with higher demand and consequently create local congestion in some downstream junctions with short links. Another consequence is that a larger hysteresis and more scatter can be observed compared to the NFD for the PC scenario, suggesting less homogeneous distribution of density in the PN.

Interestingly, the top of the NFD in the PCQ scenario seems slightly sharper than in the NPC and PC scenarios, i.e., PCQ allows for higher accumulations and the same throughput in this simulation setup for some replications.

2) Analysis of relative queues: For the evaluation of the relative queues, three replications, one for each scenario, with closest delay values to the respective average network delay have been chosen. The time series of relative queues at the gated links are shown in Fig. 4. The vertical dashed lines indicated the periods of activation of perimeter control.

In Fig. 4(a) it can be seen that in the NPC scenario queues grow strongly at some of the gated links and last for about two hours. In particular, at gated link 2 and, for short periods, at gated link 5, the queue lengths exceed the stipulated maxima. Thus, even without perimeter control, queues are formed due to a combination of spill back effects from inside the network and excessive demand. On the other hand, on gated links with low demand, for instance gated links 1, 6 and 8, queues are hardly noticeable for most of the time.

When applying perimeter control without queue balancing (PC) (Fig. 4(b)), the inflow is restricted and queuing...
starts roughly, simultaneously at all gated links. However, the queues last about half an hour less than in the NPC case. Despite the improvement due to perimeter control, the relative queues remain unbalanced and the maximum values are slightly exceeded for short periods in gated links 2 and 5.

The comparison between Fig. 4(a) and (b) resolves an important general concern regarding perimeter control: PC does not necessarily lead to higher queues than in the NPC case. As already briefly discussed in the introduction, throughput is higher with PC whereas with NPC links within the PN may be over-saturated and backspiling. Consequently, lower flow can be served at the boundary links and longer queues may be experienced with NPC than with PC.

The results of the application of perimeter control with queue balancing (PCQ) can be seen in Fig. 4(c). Although the relative queue lengths are not perfectly equal, they are more similar than without the queue balancing. The duration of the queues is even shorter in this scenario.

As noted in Section III-B, some of the unbalanced queue values in the PCQ scenario may be the result of discrepancies between ordered flow at the gated links and the realized flow due to saturation. Indeed, Fig. 5(a) with the average of ordered and actual flows at the eight gated links in the PCQ scenario during the period in which perimeter control is activated, shows considerable difference between the two values at some gated links. In Fig. 5(b) the time series of ordered and actual flow at gated link 3 in the PCQ scenario are shown. The vertical dashed lines indicate the activation and deactivation times of perimeter control. We see that the actual flow follows closely the ordered flow from activation until $t = 2.2$ h. Thereafter, the ordered flow increases and the actual flow departs. In this case, there is a lack of demand and the difference between the two values is not harmful.

In Fig. 5(c), on the other hand, we see that for gated link 7 the ordered flow reaches the lower bound (saturates) as soon as perimeter control is activated. Nevertheless, in the first 20 minutes of control the actual flows remain below the lower bound most of the time. This is the result of a localized congestion downstream of gated link 7 combined with a high demand and reflects in an increasing queue in that link (Fig. 4(c)) that cannot be balanced with the others.

Localized congestion within the PN may be avoided with the use of adaptive traffic control strategies. As a matter of fact, the effectiveness of perimeter control may be enhanced if combined with adaptive control as demonstrated by [21].

3) Analysis of delay: Similar to what was verified with queues in Fig. 4, experienced delays at the gated links are higher and last longer in the NPC scenario (not shown). Delays with PCQ (Fig. 6) are also lower than with PC (not shown). We note from the figure, however, that balanced queues do not necessarily reflect into balanced delays, which may be perceived unfair by the drivers.

V. CONCLUSIONS

A knapsack-based optimization flow distribution aiming at balancing relative queues at gated links has been proposed for the urban traffic flow perimeter control. In a comparative simulation study, queuing at the gated links have been analyzed thoroughly. The study has interestingly revealed that implementing perimeter control may not necessarily result in higher queue length at the gated links than the no perimeter control case. Applying the proposed queue management approach reduced the queuing period at the gated
links compared to the case without queue management. The corresponding queue propagation (spill-back) to the upstream junctions reduced remarkably. This is indeed an important finding towards the real-field implementation of perimeter control strategies, assuring the efficiency of gating without negatively affecting the vicinity of the gated links. However, the proposed approach does not guarantee balanced waiting times at the gated links. Thus, from equity and fairness point of view, this might be an issue. The comparison of queue-length-based and waiting-time-based flow distribution approaches is the subject of an on-going research.

ACKNOWLEDGMENT

The support of TSS - Transport Simulation Systems is acknowledged for providing the AIMSUN license.

REFERENCES


