Traffic-responsive signals combined with perimeter control: investigating the benefits

Mehdi Keyvan-Ekbatani[a*], Xueyu (Shirley) Gao[b], Vikash V. Gayah[b], Victor L. Knoop[c]

[a] *Civil and Natural Resources Engineering, University of Canterbury, New Zealand*
[b] *Department of Civil and Environmental Engineering, The Pennsylvania State University, PA, United States*
[c] *Department of Transport and Planning, Delft University of Technology, Delft, The Netherlands*

ABSTRACT

Empirical and research evidence suggests that traffic-responsive signal control strategies are generally not as efficient in over-saturated traffic conditions characterized by queue spillbacks. Recent studies on (MFD/NFD) have identified a destabilizing gridlock process that arises in congested networks which precludes efficient operation. The MFD has also been used to develop urban traffic control solutions (e.g. perimeter or gating control) to avoid congestion. The purpose of this paper is to explore the benefits of combining gating with locally adaptive traffic signals through micro-simulation of the Chania, Greece traffic network. Two adaptive traffic signal strategies are considered with the perimeter control strategy. The results of the combined gating/adaptive signal control scheme are compared to gating under fixed traffic signals and the implementation of adaptive signals only. The convincing outcome of this simulation study motivates the real-field implementation of gating/perimeter control in the cities, without even changing the existing adaptive control strategies.

Keywords: Macroscopic or Network Fundamental Diagram, Gating or Perimeter Control, Traffic-Responsive Signal Control, Large-Scale Networks

1. Introduction and background

As the world population continues to grow and people move towards denser urban environments, urban traffic congestion will continue to be a significant problem with large economic and environmental consequences (Schrank et al., 2012). Strategies to combat urban traffic congestion mainly focus on either travel demand management or the implementation of traffic control to improve the capacity for vehicle movement. Demand management strategies include congestion pricing (Verhoef, 2002), carpooling (Correia and Viegas, 2011), staggered work hours and transit improvements. The goal of these strategies are to reduce congestion by reducing vehicle-trips.

On the other hand, traffic control strategies in urban environments seek to maximize the capacity for vehicle movement. The majority of these efforts are spent on the careful selection of traffic signal timings, since traffic signals serve as the most restrictive and frequent bottleneck in urban environments. In the most simplistic case, fixed signal timings are selected based on prevailing traffic conditions. These can be constant throughout a day or, more likely, changed at regular intervals that correspond to different demand patterns (e.g., peak hours). Complex mathematical frameworks have also been proposed to optimize traffic signal timings in real-time based on current traffic measurements. The most comprehensive of these frameworks are centralized control strategies in which all signal timings are jointly optimized to maximize traffic performance and minimize congestion on a network. Examples include OPAC (Gartner, 1983),

* Corresponding author.
E-mail addresses: mehdi.ekbatani@canterbury.ac.nz (M. Keyvan-Ekbatani), xug103@psu.edu (X. Gao), gayah@engr.psu.edu (V.V. Gayah), v.l.knoop@tudelft.nl (V.L. Knoop)

PRODYN (Farges et al., 1983) and other optimization frameworks recently proposed in the literature (Han et al., 2015; Lo, 1999a, b; Ukkusuri et al., 2013). The computational complexity of these algorithms increases exponentially with the network size (Knoop et al, 2013). Consequently, these centralized strategies are not suited for application in larger networks or regions in practice.

Adaptive or traffic responsive signal control strategies lie somewhere within this spectrum. These strategies alter signal timings in real-time based on traffic data, but do so in a decentralized fashion. Thus, signal timings at a signal location will be adjusted based on traffic data from detectors immediately upstream or downstream. In some cases, constraints will be used to facilitate coordination between adjacent signals so that network-wide effects can be considered. Common examples of adaptive signal control schemes used in practice include SCATS (Lowrie, 1982; Slavin et al., 2013) and SCOOT (Hunt et al., 1982). Another practical approach that is gaining traction for real-world application is TUC (Diakaki et al., 2002). Unfortunately, most existing adaptive traffic control strategies face limitations when applied to oversaturated or congested traffic conditions that are frequently observed in large-scale urban cities. In fact, a recent study used a simple analytical model to show that typical locally adaptive signal control schemes tend to have little to no effect on traffic at network level when congested due to downstream congestion and queue spillbacks (Gayah et al., 2014). Moreover, the aforementioned methods may allow too much traffic to enter from the boundary of a network, particularly if the boundary is less congested, which can exacerbate queue spillbacks in the congested areas. Adaptive signals also tend to act only after congested begins to occur. These could be important reasons why existing adaptive traffic signal control strategies do not operate efficiently in highly congested urban road networks. These could be important reasons why existing adaptive traffic signal control strategies do not operate efficiently in highly congested urban road networks.

Recent works have shown that urban traffic networks might be more efficiently controlled by limiting the rate at which vehicles are allowed to enter to avoid oversaturated conditions altogether within the busiest parts of a network. This perimeter gating strategy relies on relationships between traffic variables measured network-wide, which have been the subject of research for several decades (Edie, 1965; Godfrey, 1969; Mahmassani et al., 1987; Mahmassani et al., 1984; Olszewski et al., 1995; Smeed, 1966; Zahavi, 1972). (Daganzo, 2007) used a hypothetical relationship between network accumulation—how many vehicles are within a network—and production—how many vehicles are able to be served—at a given time to first show that allowing a network to become congested can trigger a devastating gridlock process that reduces efficiency as vehicles continue to enter. This work also showed how gating could help improve efficiency and reduce delays during the entire peak period, even for vehicles that had to wait to enter the network. The relationship this work was built upon—known more commonly as the Macroscopic or Network Fundamental Diagram (MFD or NFD)—was verified using empirical data first by (Geroliminis and Daganzo, 2008) and later by others (Buisson and Ladier, 2009; Tsubota et al., 2014). Since these initial efforts, many studies have examined various features of MFDs (Gayah and Daganzo, 2011a; Gayah and Daganzo, 2011b; Saberi et al., 2014; Knoop et al., 2015; Leclercq and Geroliminis, 2013; Leclercq et al., 2015; Mahmassani et al., 2013) and how they might be estimated using a variety of data sources (Du et al., 2015; Gayah and Dixit, 2013; Leclercq et al., 2014; Nagle and Gayah, 2014). Perhaps more importantly, several studies have developed real-time control strategies using the MFD to limit vehicle entries within a protected network (PN) using traffic signals to ensure that it does not become oversaturated (Haddad et al., 2013; Geroliminis et al., 2013; Haddad and Shraiber, 2014; Haddad and Zheng, 2018; Keyvan-Ekbatani et al., 2012; Keyvan-Ekbatani et al., 2015; Keyvan-Ekbatani et al., 2013; Keyvan-Ekbatani et al., 2014; Ramezani et al., 2015; Yildirimoglu et al., 2015). Recent works extend the single-mode MFD to a bi-modal where not only the vehicular dynamics but also passenger flow dynamics has been taken into account (Zheng and Geroliminis, 2013; Geroliminis et al., 2014). Using the notion of MFD, Ramezani and Nourinejad (2017) proposed a model predictive control (MPC) approach to control the taxi dispatch system. Recenlty, Keyvan-Ekbatani et al. (2016b) propsed a perimter control strategy which incorporates the queuing and waiting time at the gated links. These studies all verify the original findings of (Daganzo,

2007) that networks should be controlled at their critical accumulation, which is the point on the MFD that offers the highest production rate. However, the integration of this perimeter control strategy with traditional adaptive signal control schemes has yet to be considered even though such a combination might offer significant benefits. For one, avoiding congestion through perimeter flow control allows the traffic responsive signal control strategies to work in situations they perform best. Furthermore, adaptive signal control has shown to increase a network's MFD (Gayah et al., 2014), which can improve the efficiency of the perimeter flow control. The preliminary results of our work has been presented in (Keyvan-Ekbatani et al., 2016a). Recently, some attempts were carried out in this direction which are different compared to what we have studied here. For instance Yang et al. (2017) formulated a dynamic system model, and designed a Model Predictive Control (MPC)-based controller which optimizes the performance at the local and the network level as a whole in connected system environment. Mehr et al. (2017) proposed a control algorithm which uses alternating direction method of multipliers without applying MFD concept and was tested on a small network with only 6 intersections and 16 links. In our work, we used widely implemented traffic strategies with a simple but efficient perimeter control algorithm which is not depending on demand predictions and accurate modelling procedure. Other works like Zargari et al. (2018) used heuristic method for a simple simulation study to have a link-based metering.

In light of this, the objective of this paper is to examine the combined impacts of perimeter flow control and traffic responsive signal control on overall network efficiency using micro-simulation (see Fig.1). Two adaptive signal control strategies are considered. The first is a simple volume-based strategy in which traffic signal timings are continuously updated every cycle based on volume counts from upstream detectors. The second adaptive signal control strategy is a modified version of the realistic SCATS algorithm, which appears to be one of the most applied strategies in practice currently. The perimeter control strategy applied in this paper is a feedback-based strategy that continuously updates green times on the periphery of the PN based on the accumulation inside. Previous work has shown that adaptive signal control can have positive effects on the MFD when the network is moderately saturated by helping to provide more homogeneous vehicle distributions across the network (Gayah et al., 2014; Zhang et al., 2013). These positive effects manifest as improved capacities and slightly higher critical accumulations on the MFD, both of which are vital parameters in the implementation of gating. Thus, adaptive signal controls should provide networks with higher peak efficiencies that can be leveraged during the gating process.

Our results reveal that perimeter control provides more drastic improvements in network efficiency—measured in average vehicular delay and mean speed within the network—when compared to adaptive signal control only. This is not unexpected as the adaptive signal control fails to prevent congestion within the network, whereas perimeter control can keep the network operating at peak efficiency by holding some vehicles outside of the PN. More importantly, the combination of the two strategies appears to yield significant benefits over gating alone. The increase in network capacity and associated critical accumulation provided by adaptive signal control results in more vehicles being able to enter the network at any point in time and thus shorter queues at the boundary of the PN when gated. This is an important finding for gating of multi-region networks where queues at the boundary of one network can negatively impact operations in adjacent networks. Furthermore, the combination of both strategies results in more homogeneous traffic conditions within the PN network, which results in more efficient network operation. Overall, the study finds that travel delays and congestion can be considerably reduced with the combined strategy. In addition, it has been observed that the virtual queue length (i.e. mimicking the queue spillback to the vicinity area) reduced remarkably, when traffic-responsive signal control applied in the sub-network (or protected region).

The rest of this paper is organized as follows. Section 2 discusses the traffic-responsive signal control strategies which were applied in this study. The MFD-based perimeter control strategy is presented in section 3. The simulation set-up and test-bed description are discussed in section 4. The comparison of the network performance under traffic-response signal control and perimeter traffic control has been carried out in section 5. Section 6 examines the network efficiency in case of a hierarchical urban traffic control

(i.e. traffic-responsive control at the lower and perimeter control at the higher level). This is followed by a discussion and conclusion section at the end.

2. Implemented Traffic-Responsive Control Strategies

Adaptive traffic signal systems modify signal timings in real-time based on prevailing (local) traffic conditions. A potential challenge of most of the existing traffic-responsive strategies is that they are mainly effective in only one or two of the traffic regimes (i.e. under-saturated, saturated and over-saturated; see Van de Weg et al. 2016). In this paper, two different adaptive traffic responsive strategies of varying degrees of complexity were considered: (1) a volume-based approach; and, (2) a realistic adaptive signal strategy that is widely implemented throughout the world, SCATS. More advanced traffic-responsive control strategies exist in the literature that could be applied to this study. However, these two strategies were chosen due to their simplicity, practicability and the fact that both emulate the primary purpose of any traffic-responsive signal control system: to provide more green time to the approach(es) with the higher current traffic volumes. Furthermore, recent theoretical and simulation-based studies (Zhang et al, 2013; Gayah et al, 2014) have used these specific strategies to unveil the impacts of adaptive signal control on overall network performance. These studies find that both strategies improve network conditions (demonstrated by larger MFDs and higher travel speeds for a given network density) during lightly and moderately congested conditions. However, these strategies generally fail to improve network performance under heavy congestion. The remainder of this section describes each of these strategies in more detail.

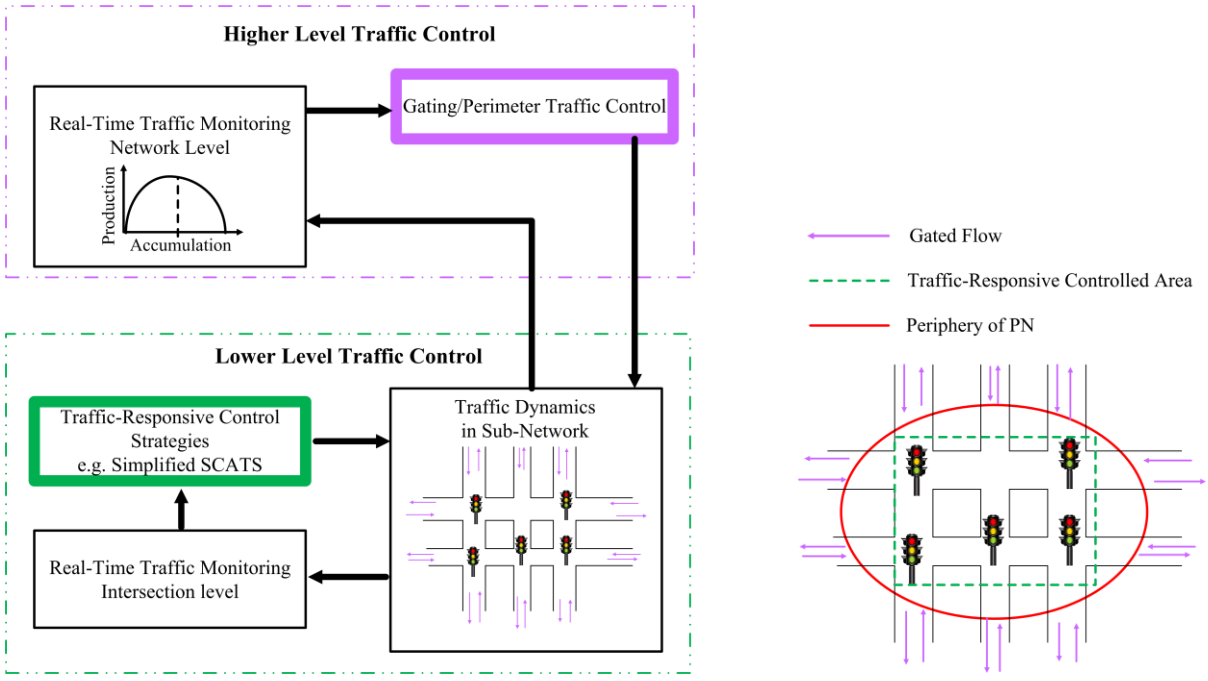

Figure 1 Structure of the proposed multi-layer urban traffic control (left); schematic illustration of the combined traffic control strategy in the network (right)

2.1. Volume-based (VL) approach

In the first strategy, the same fixed cycle length of 90 seconds was adopted at all signalized intersections. The available green time for movement was divided upon competing approaches based on the traffic demand on each respective approach. A simple proportional algorithm was used in which the green time was allocated proportional to flows measured during the previous cycle at detectors on the

upstream approaches. Specifically, the green time on subject approach i during cycle $k = 0,1,2,...$, $g_i(k)$, is determined as follows:

$$g_i(k) = (C - L - G_{\min}).\frac{v_i(k-1)}{\sum_i v_i(k-1)} + g_{i,\min} \quad (1)$$

where C is the fixed cycle length, L is the lost time for vehicle movement (usually due to and directly proportional to the number of phase changes), $g_{i,\min} = (1-\alpha)\cdot g_{i,\text{fix}}$ (i.e. the minimum green time on approach i), α is the percentage of cycle length which could be adjusted based on the traffic volume data (i.e. 60% in this paper), $g_{i,\text{fix}}$ is the fixed green time for approach i, $G_{\min} = \sum g_{i,\min}$ and $v_i(k\text{-}1)$ is the volume observed on approach i during cycle $k-1$. Thus, this strategy is called the *volume-based* strategy.

All available green time was allocated in this way. Thus, it is possible that some approaches receive zero green time if no vehicles are queued at the approach. Although this might not be realistic, it provides the maximum flexibility and offers the largest green times to the major direction(s). It was first applied to examine the impacts of adaptive signal control on the MFD in an analytical investigation (Gayah et al., 2014) and has been used here to represent the simplest traffic-responsive control algorithm that might be considered in practice.

2.2. Modified SCATS

The second strategy implemented in this study is a simplified version of the Sydney Coordinated Adaptive Traffic System SCATS. SCATS is currently applied in many cities throughout the world to alter signal timings based on prevailing traffic conditions. The version of SCATS implemented in this study has been adopted from (Zhang et al., 2013) where it was used to assess the impacts of adaptive signal control on the MFD using simulation.

In this modified SCATS strategy, both the green time and total cycle lengths are variable and adjusted based on volume data obtained from upstream loop detectors. As described in (Zhang et al., 2013), an appropriate cycle length is first selected based on the volume ratio observed during the previous cycle $R(k-1)$ which is derived from:

$$R(k\text{-}1) = \max_i \left[\frac{v_i(k-1)}{g_i(k-1)\cdot v_{i,\max}} \right] \quad (2)$$

where $v_{i,\max}$ is the capacity of approach i. The cycle length is designed to maintain a volume ratio between 0.85 to 0.95 during the next cycle and is selected based on the following rules:

$$C(k) = \begin{cases} min\left[C(k-1) + \Psi,\ C_{\max}\right] & \text{if } R(k-1) > 0.95 \\ max\left[C(k-1) - \Psi,\ C_{\min}\right] & \text{if } R(k-1) < 0.85 \\ C(k-1) & \text{otherwise} \end{cases} \quad (3)$$

where $C_{\min}$ and $C_{\max}$ are the minimum and maximum cycle lengths, respectively, Ψ is the fixed amount of increment/decrement allows for the change of cycle length , and $R(k-1)$ represents the volume ratio at a given intersection during cycle $k-1$. The cycle length is allocated among the competing approaches based on the vehicle demand on each approach. The following equation is used to allocate this green time:

$$g_i(k) = \left(C(k) - L - G'_{\min}\right).\frac{d_i(k-1)}{\sum_i d_i(k-1)} + g'_{i,\min} \quad (4)$$

where $G'_{\min}$ is the sum of the minimum green times for each approach, $g'_{i,\min}$ is the minimum green time allocated to each approach, and $d_i(k-1)$ is the vehicle demand on approach i. For this paper, the following values were used for the adaptive signals with SCATS: $C_{\min} = 60\,\text{s}$, $C_{\max} = 132\,\text{s}$, $\Psi = 6\,\text{s}$ and $g'_{i,\min} = 6\,\text{s}$. The modified SCATS strategy is similar to the volume-based strategy described in Section 2.1 in that available green time is divided among competing approaches based on observed traffic volumes. However, the SCATS is more realistic as it considers the implementation of minimum green times (to provide a minimum level of service for minor approaches) and flexible cycle lengths based on the observed volume-to-capacity ratio.

3. Implemented Perimeter Control Strategy: A Brief Review

This section provides an overview of the notion of Macroscopic Fundamental Diagram (MFD), the applied MFD-based model and the implemented feedback controller (Keyvan-Ekbatani et al., 2012).

3.1. MFD concept

Recently, it was found that the notion of a fundamental diagram (e.g. in the form of a flow-density curve) can be applied (under certain conditions) to two-dimensional, spatially distributed urban road networks. The two-dimensional fundamental diagram relates the summation of weighted flow and average density. A unimodal curve is generally expected: as the accumulation or density is increasing, the traffic flow increases up to a point that the capacity is reached. Exceeding the range with critical density, the network starts degrading and enters the over-saturated region. The aim of most network-wide traffic control strategies is to maintain the overall traffic state in the range of maximum total flow and to avoid spillover and gridlock creation by applying various traffic management tools (e.g. traffic signal optimization, gating, route guidance and etc.). According to the definitions introduced in (Edie, 1965; Mahmassani et al., 1984; Keyvan-Ekbatani et al., 2012), the MFD can be also plotted as Total Travelled Distance (*TTD* in veh·km per h) for the y-axis and Total Time Spent (*TTS* in veh·h per h) for the x-axis. The *TTD* and *TTS* are obtained from the (emulated) loop measurements via the following equations:

$$TTS(k) = \sum_{z\in\mathbb{M}} \frac{T\cdot\hat{N}_z(k)}{T} = \sum_{z\in\mathbb{M}} \hat{N}_z(k) = \hat{N}(k) \quad (5)$$

$$TTD(k) = \sum_{z\in\mathbb{M}} \frac{T\cdot q_z(k)\cdot L_z}{T} = \sum_{z\in\mathbb{M}} q_z(k)\cdot L_z \quad (6)$$

where *z* is the link where a measurement is collected; $\mathbb{M}$ is the set of measurement links; k 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*; $\hat{N}(k)$ is the estimated number of vehicles within the links in the set of $\mathbb{M}$. *TTS* is the estimated total time spent, which turns out to equal the estimated number of vehicles in the subset of the links 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:

$$\hat{N}_z(k) = L_z\cdot\frac{\mu_z}{100\lambda}\cdot o_z(k-1) \quad (7)$$

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. The link estimated number of vehicle is fairly precise, if the detector is located around the middle of the link.

3.2. MFD-based Modeling

The dynamics within the PN might simply be modelled by applying the notion of MFD. According to the conservation equation for vehicles in the PN we have (Keyvan-Ekbatani et al., 2012)

$$\dot{N}(t) = q_g(t) + q_d(t) - q_{out}(t) \tag{8}$$

where q_g is the gated flow, q_d represents other (non-gated or internal) inflows to the PN (disturbances), q_{out} the flow which exits the PN and N is the number vehicles in the PN (see Figure 2 for the schematic illustration). Applying the modelling steps in (Keyvan-Ekbatani et al., 2012; Keyvan-Ekabatani et al., 2015), the discretized linear model can be derived as

$$\Delta TTS(k+1) = \mu \cdot \Delta TTS(k) + \zeta \cdot \Delta q_g(k) \tag{9}$$

where Δ-quantities reflect corresponding deviations from nominal (or the desired steady values). Applying a model identification approach, the model's parameters μ and ζ might be derived utilizing the available data for $TTS(k)$ and Δq_g (either simulation or real data).

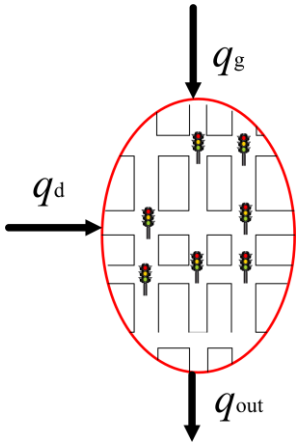

Figure 2 Schematic illustration of the single-region perimeter control (adopted from Keyvan-Ekbatani et al., 2012)

3.3. Feedback-Based Control

The following proportional-integral-type (PI) feedback regulator ((Keyvan-Ekbatani et al., 2012)) has been applied to keep the system state, $TTS(k)$, at the steady level $\hat{TTS}$

$$q_{\mathrm{g}}(k) = q_{\mathrm{g}}(k-1) - K_{\mathrm{P}}\left[TTS(k) - TTS(k-1)\right] + K_{\mathrm{I}}\left[\hat{TTS} - TTS(k)\right] \quad (10)$$

K_{P} and K_{I} are the proportional and integral gains, respectively. Proper controller parameters may be derived by the methodology presented in (Keyvan-Ekbatani et al., 2015b) or manual fine tuning. The ordered flow g_{g} by (10) might be distributed among the gated links $g_{\mathrm{g},i}$ based on different policies and translated to green phases in the signal control plans of the intersections. Note that, in this paper, the gated flow distribution is according to the saturation flow of the respective links s_i and the spillback recognition algorithm developed in (Keyvan-Ekbatani et al., 2015a).

3.4. Converting the Distributed Flows into Green Phases

To execute the gating strategy at the gated junctions, the distributed ordered flow should be translated into green phases. This is done by

$$g_{\mathrm{g},i} = \frac{q_{\mathrm{g},i} \cdot C}{s_i} \quad (11)$$

where C and $g_{\mathrm{g},i}$ are the traffic light cycle length, and gated green time, respectively. The fixed-time signal plan of the considered urban network is modified at the gating junctions when real-time gating actions are present, by modifying accordingly the involved signal stage (phase) 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.

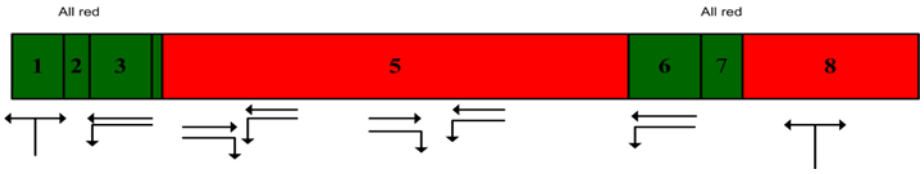

Figure 3 Signal plan of the intersection at the easternmost side of the PN in Figure 4

Figure 3 exhibits the fixed-time signal plan of the traffic light at the easternmost side of the perimeter of PN. The *interphases* are shown by dark green along with the corresponding movements shown by arrows which remain constant during the control period. *Interphases 2* and *7* are all-red stages in which all movements at the junction receive red light. During gating action, the movement into the PN should be metered which belongs to *stage 5*. Thus, the gated signal stage (phase) is *stage 5* and any positive or negative deviation of *stage 5* during the gating activation, is assigned to *stage 8*. By this, the cycle length is kept constant during the gating period. Generally, the stage re-arranging should be such that $g_{\min,i} < g_{\mathrm{g},i} < g_{\max,i}$ where $g_{\max,i} = \sigma_{\text{gated},i} + \sigma_{\text{non-gated},i} - g_{\min,i}$ and σ is the stage duration. In case of existence of more than 2 stages, the positive or negative deviation of the gated stage either can be assigned to the stage which involves the movement of exiting flow from the PN or can be assigned proportionally to the other stages based on $\bar{g}_i$ (nominal green phases).

If a gated link (whose outflow enters the PN) receives its right-of-way in the same stage as an opposed-direction link (whose outflow exits the PN, e.g. in *stage 5* in Figure 3), an appropriate re-staging has been introduced to avoid unwanted and unnecessary delays at PN-exiting links during the gating activation. More specifically, an extra 0s stage is defined in the signal plan of that junction. This additional stage will be active during the gating period (in the signal cycle) in order to serve the green phase dedicated to the outflow

exiting the PN. This phase separation is carried out as follows:

$$g_{g,i} < \bar{g}_i \Rightarrow \begin{cases} \sigma_{\text{gated},i} = g_{g,i} \\ \sigma_{\text{added},i} = \bar{g}_i - g_{g,i} \\ \sigma_{\text{non-gated},i} = g_{\max,i} + g_{\min,i} - \bar{g}_i \end{cases} \quad (12)$$

$$\bar{g}_i \le g_{g,i} \Rightarrow \begin{cases} \sigma_{\text{gated},i} = g_{g,i} \\ \sigma_{\text{added},i} = 0 \\ \sigma_{\text{non-gated},i} = g_{\max,i} + g_{\min,i} - g_{g,i} \end{cases} \quad (13)$$

As it can be seen from (13), the gated green (i.e. ordered by the controller) might be even larger than the nominal (or original green) of the gated signalized links during gating activation, if the estimated number of vehicles in the PN is below the critical accumulation.

4. Test-Bed Description and Simulation Setup

The urban surface street network of Chania, Greece has been applied as a test-bed to implement the control strategies introduced in this study; see Figure 4 which depicts the portion of the Chania network modelled in AIMSUN (TSS, 2016). The central business district (CBD) of the Chania urban road network, where the congestion usually starts during the peak period, is considered as the PN (the shaded area in Figure 4). However, a few slight modifications have been made to facilitate this analysis. The majority of intersections within the PN are unsignalized, which does not offer an opportunity to examine the integration of traffic-responsive signal control and perimeter flow control. To alleviate this issue, some of these unsignalized junctions within the PN were signalized in the simulation to make the implementation of the adaptive traffic control strategies more feasible without changing the geometric and topographic characteristics of the network.

It should be noted that signals are not coordinated in this network. Instead, adaptive control is applied without any explicit coordination between adjacent signals. And although the adaptive signal might seem to be implemented only along one corridor, they are actually implemented dimensionally in the network (e.g., see the right hand side of Figure 4). Furthermore, as previously mentioned, the impacts of these adaptive control (along) were studied in (Gayah et al, 2014) and found to be highly beneficial. Lastly, we would like to note a previous study that specifically studied in the impacts of signal coordination (alone, without adaptive signal control) on the MFD (Girault et al, 2016). This work shows that coordinate on a two-dimensional grid network did not significantly impact the MFD due to the balance between gains and losses of coordination of vehicles traveling in multiple directions and dimensions.

Additionally, existing signalized intersections at the periphery of the network were considered as gating junctions and used to regulate the flow of vehicles into the PN. These gating junctions were selected based on their location and prevailing demand patterns (i.e., they represented the junction at which most of the traffic enters the PN from outside) as well as the availability of space to store vehicle queues at the boundary. The location of the eight gated links at the periphery of the PN has been indicated by red arrows in Figure 4 illustrating the queuing locations upstream of the gating intersections.

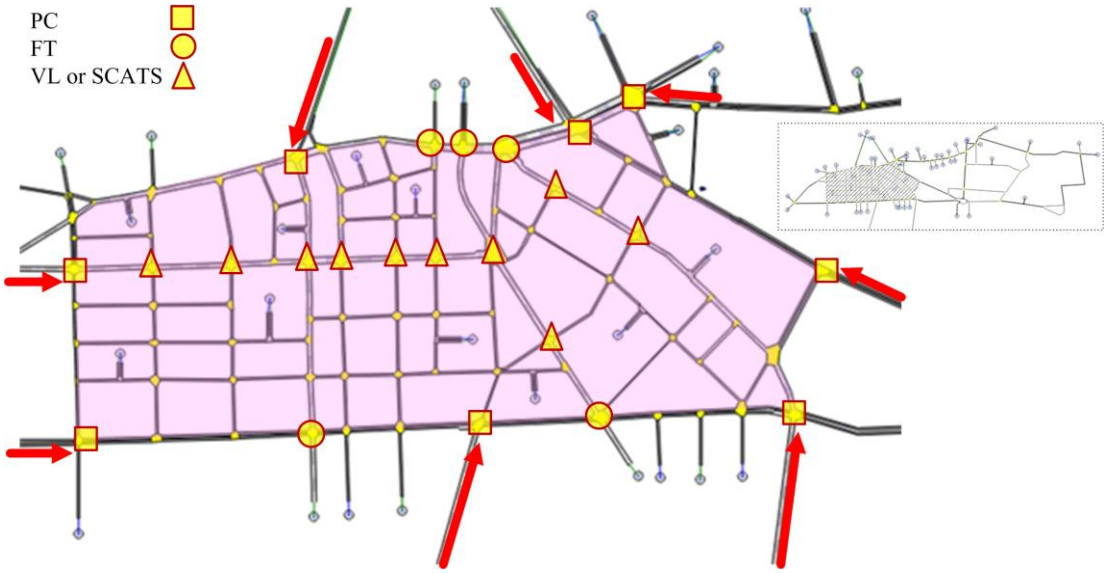

Figure 4 Chania (Greece) urban network modelled in AIMSUN; indication of traffic lights applied for the control strategies within the PN (left) and the illustration of the larger part of the network (right)

In the middle of each link inside the PN, a loop detector has been inserted and used to collect relevant traffic measurements in the simulation at regular 90-s intervals. Realistic but not exact O-D flows have been taken into account (with regard to the utilized O-D rates). To mimic the peak and off-peak period, as illustrated in Figure 5, a 4-hour (7:00 AM-11:00 AM) trapezoidal demand profile has been introduced to the microscopic simulator AIMSUN. After 8:20 AM, the demand is gradually decreased in a very short time intervals (i.e. 5-minute intervals) to decrease the potential scatter in the MFD. The total number of vehicles served for all simulation scenario with different random seeds are quite similar, since the network has been emptied at the end of the 4-hour simulation. A c-logit-based route choice system distributes the cars over the different paths from each origin to each destination, while running the simulation (Dynamic Traffic Assignment enabled). Ten different replications (i.e. simulation runs) for each investigated scenario have been carried out to account for the stochasticity of the microscopic simulation AIMSUN. Three network-wide performance indexes have been taken into account to evaluate efficiency of the various traffic control strategies: the average vehicle delay (s/km), average virtual queue (veh) and total time spent (h).

5. Simulation Scenarios

Multiple scenarios are considered to compare network performance under the individual and combined urban traffic control strategies described in Sections 2 and 3. The following combinations are considered (NPC and PC stand for No-Perimeter Control and Perimeter Control, respectively):

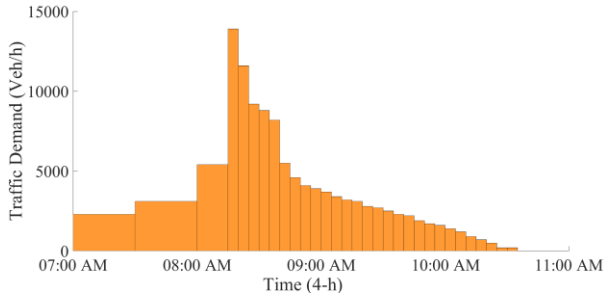

Figure 5 Demand profile applied for the 4-h simulation (7:00-11:00 AM)

- ***NPC (FT)***: A Fixed-Time (FT) traffic control has been utilized for the signalized junctions within the PN. The usual peak-hour FT signal plan of the Chania urban road network has been applied in the simulations (while optimized existing fixed signal timings are considered for the fictitious signals added for this study). The utilized traffic lights for this scenario are indicated by triangle, square and circle symbols in Figure 4.
- ***NPC (VL)***: The Volume-Based (VL) traffic-responsive signal control plan described in Section 2.1 is applied to the internal traffic signals within the PN (illustrated by the triangle in Figure 4). The remaining signals on the periphery of the PN maintain a fixed-time plan.
- ***NPC (SCATS)***: In this scenario, the modified SCATS traffic-responsive control strategy described in Section 2.2 is applied to the internal traffic signals within the PN (illustrated by the triangle in Figure 4). The remaining signals on the periphery of the PN maintain a fixed-time plan.
- ***PC + FT***: In this scenario, traffic lights at the periphery of the PN (indicated by square symbols) are controlled by the MFD-based control strategy. The remaining traffic lights are operating according to a fixed-time (FT) signal plan. The red arrows in Figure 4 indicate gated inflow direction into the defined PN.
- ***PC + VL*:** This simulation scenario combines the volume-based (VL) traffic-responsive control strategy described in Section 2.1 with the MFD-based gating. The traffic lights indicated by a triangle in Figure 4 are controlled using the VL control strategy. The gated junctions (square symbols in Figure 4) are controlled using the perimeter flow control algorithm. The remaining peripheral signals (circles in Figure 4) operate using the fixed-time plan.
- ***PC + SCATS***: This scenario is identical to the previous scenario except the modified SCATS signal control system is used to control the internal signalized junctions.

The NPC (FT) has been chosen as the base simulation scenario with which to compare the performance of the remaining scenarios and demonstrate the benefits of the combined traffic-responsive signal control and gating strategies.

6. Simulation Results

The traffic-responsive and the MFD-based gating control strategies introduced in Sections 2 and 3 were implemented in the simulation environment described in Section 4. This section first provides a summary of the principal findings of this work. Then, it provides an examination of the individual simulation results to explain these findings in more detail.

6.1. Main findings

Figure 6 provides a comparison of flow in the PN over time for four of the simulated scenarios. The median flow values for the 10 simulated iterations for each of the scenarios are presented. This figure demonstrates a few of the most pertinent findings of the present study. Note that flows are almost identical among the strategies for the beginning of the simulation period (up until $t = 1.5$ h). This is reasonable since the control strategies considered will not significantly impact operations when the network is very lightly congested. For the perimeter gating control, changes are only implemented when the network exceeds the

critical accumulation values associated with congestion. For the traffic-responsive signal control strategies, signal timings only change when flows become unbalanced among competing approaches as the network becomes more congested. Due to local gridlock and spillback within the PN during the peak period, traffic performance within the PN starts degrading and the flow drops to 250 veh/h in NPC (SCATS). This finding is in line with the literature (Denney Jr. et al., 2008; Aboudolas et al., 2009) as indicates that the performance of SCATS and most of the existing practicable traffic control strategies may deteriorate when severe congestion persists during the peak period; i.e., during the period from $t \in [1.6, 2.7]$ h.

Figure 6 reveals the following major findings:

1. When perimeter control is applied (PC strategies), congestion in the network dissolves much earlier (at around $t = 2.7$ h instead of $t = 3.3$ h and $t = 3.7$ h) compared to the non-perimeter gating control cases (NPC strategies).
2. The flow capacity of the protected network increases when traffic-responsive signal control strategies are applied (SCATS strategies) compared to when they are not applied (FT strategies).
3. The combination of perimeter gating control and traffic-responsive signal control strategies (PC + SCATS) allows higher flow capacities to be maintained compared to traffic-responsive signal control strategies alone (NPC (SCATS)) or perimeter gating strategies alone (PC + FT).
4. Congestion dissolves much faster in the MFD-based perimeter gating cases (PC strategies) compared to non-perimeter gating cases (NPC strategies). For instance, at $t = 3.3$ the average PN flow decreased from around 300 veh/h to 80 veh/h. In other words, the network has been emptied much faster in PC strategies than the NPC strategies.

6.2. MFD best-fit approach

For an on-line traffic monitoring, the MFDs of the shaded area in Figure 4 for the simulated scenarios have been derived based on the *TTS* and *TTD* calculated by (5) and (6), respectively. Using the generalized definitions of Edie (Edie, 1965), the network average flow and average density have been obtained (for the y- and x-axis, respectively) by dividing the *TTS* and *TTD* values by the total PN's length (12.44 km). A 90 s time resolution has been considered to obtain the real-time measurements to estimate the traffic states. To obtain the best fit to the MFD scatterplot in Figure 7, an approach based on Drake fundamental diagram has been utilized (Drake, 1967):

$$V = p_1 \cdot e^{-\frac{1}{2}\cdot\frac{k'}{k'_{cr}}} \tag{14}$$

where p_1 and k'_{cr} are the function parameters and k' is the density on the modified x-axis of the Drake fundamental diagram. To attain k', an asymmetry parameter (skewness) has been introduced to the function and the density k has been non-linearly transformed to:

$$k' = k^{p_2} \tag{15}$$

where p_2 is the skewness parameter. (16) might be applied to derive the k'_{cr} (i.e. $k'_{cr} = k_{cr}^{p_2}$). Applying the fundamental relationship between the speed and density, the network flow is obtained as follows:

$$Q = k' \cdot V = k^{p_2} \cdot p_1 \cdot e^{-\frac{1}{2}\cdot\frac{k'}{k'_{cr}}} = k^{p_2} \cdot p_1 \cdot e^{-\frac{1}{2}\cdot\frac{k^{p_2}}{k_{cr}{}^{p_2}}} \tag{16}$$

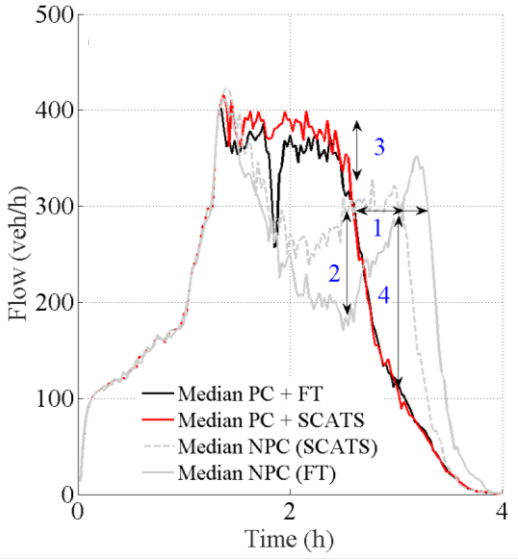

Figure 6 Main findings: (1) earlier congestion dissolution in PC +SCATS and PC +FT than the NPC (SCATS); (2) higher PN capacity in PC + SCATS than NPC (SCATS); (3) higher PN capacity in PC + SCATS than PC + FT; (4) faster congestion mitigation in PC scenarios than traffic responsive SCATS

The fitting function holds three parameters: one relates to the (free-flow) speed p_1, one to the capacity k'_{cr} and one to the skewness p_2. The parameters are tuned in order to minimize the error to the MFD curve (the Root Mean Square Error (RMSE) between predicted and actual production/flow). Note that we differentiate between the loading and the unloading phases, by considering the time instant when the density is at the maximum level.

6.3. Traffic-responsive versus fixed-time traffic control

In this subsection, the network performance under traffic-responsive control strategies (VL and SCATS) has been compared to the fixed-time (FT) scenario. In particular, the benefits of traffic-responsive strategies in terms of congestion mitigation, network capacity and critical accumulation values have been studied.

6.3.1. MFD Analysis

Figure 7 (a-c) illustrates the MFDs of the NPC scenarios (i.e. FT, VL and SCATS) for the PN loading period (i.e. approximately the first 2 hours). The red line shows the best fit to the data of the 10 different simulation runs based on the procedure outlined in Section 6.2. As displayed by the dashed horizontal line in Figure 7, the network capacity (average maximum flow) enhanced in the NPC (VL) and NPC (SCATS) scenarios (i.e. traffic-responsive control strategies) from 390 veh/h to 420 veh/h compared to when no adaptive signal control was applied. Moreover, the critical density range has been extended in case of traffic-responsive control strategy implementation form 40-44 veh/km in NPC (FT) to 44-48 veh/h in the NPC (VL) and 45-49 veh/h in the NPC (SCATS); the critical density ranges in NPC strategies has been displayed by the dashed rectangle in Figure 7 (a-c).

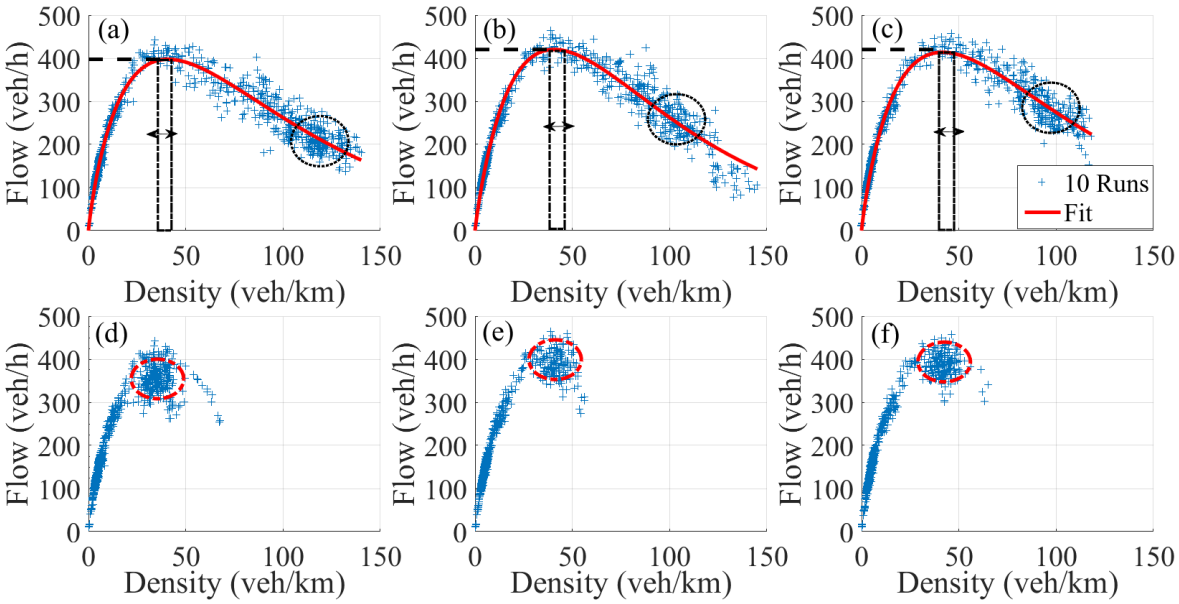

Figure 7 Loading period MFDs for the NPC (a-c) and PC strategies (d-f): (a) NPC (FT); (b) NPC (VL); (c) NPC (SCATS); (d) PC + FT; (e) PC + VL; (f) PC + SCATS

Besides the aforementioned improvements in the MFDs after implementing the traffic-responsive control strategies, the level of congestion or gridlock possibility has also been reduced. This manifests as fewer points on the congested branch and fewer high density observations in the MFDs presented in Figure 6. Observing Figure 7, the data concentration in the congested branch of the MFD (indicated by dashed black circles in Figure 7 (a-c)) is mostly in a network average density of 120-140 veh/km, 80-120 veh/km and 70-110 veh/km for the FT, VL and SCATS scenarios, respectively.

To shed more light on the network performance under traffic responsive strategies, as in Knoop et al. (2015), the traffic dynamics in the PN has been studied by analyzing the standard deviation (SD) of the link density. Figure 8 (a-c) displays the MFDs, using the median of the 10 simulation runs data, for the 4-h simulation period (i.e. loading and unloading separately) of the NPC (FT, VL and SCATS) scenarios. Note that a remarkable clock-wise hysteresis (the direction is shown by the black arrows) can be observed in all the NPC simulation scenarios.

As illustrated in Figure 8 (a-c), the network recovers at a lower average density from the congestion in the NPC (VL and SCATS) scenarios compared to the NPC (FT) scenario (at the density level of 110 veh/km instead of 140 veh/km). The color bar indicates the standard deviation (SD) of density on links within the PN. The dark red and dark blue colors indicate the highest and lowest SD values, respectively. As also observed in Knoop et al (2015), an increase of the density SD leads to a decrease in network production (or network average flow) for the same density (compare SD colors for the loading and unloading curves). For a moderately congested conditions, both traffic-responsive control strategies (VL and SCATS) seem to be beneficial to increase the network capacity compared to FT scenario and homogeneously distribute the congestion; see the loading curve in Figure 8 (b) and (c) for the density up to around 50 veh/km. In highly congested conditions, the VL and SCATS control strategies are not successfully able to mitigate the congestion and high levels of density heterogeneity are still observed in PN.

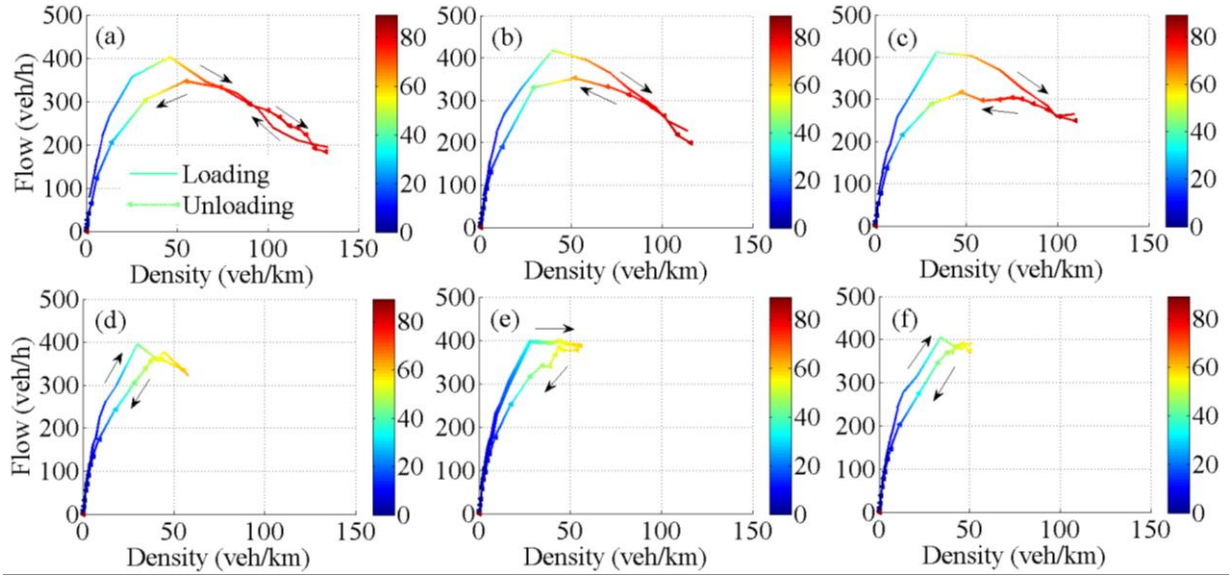

Figure 8 MFDs (median of the 10 runs) for the NPC (a-c) and PC strategies (d-f): (a) NPC (FT); (b) NPC (VL); (c) NPC (SCATS); (d) PC + FT; (e) PC + VL; (f) PC + SCATS (the link standard deviation (SD) within the PN indicated by the color-bar)

As depicted in Figure 9, both traffic-responsive control strategies are able to reduce the density SD compared to the FT simulation scenario (compare the red and blue line values with the black line for $t \in [1.5,2.2]$ h and $t \in [2.8,3.5]$ h).

6.4. Perimeter control versus fixed-time control

To implement the MFD-based perimeter control strategy, a pre-specified critical density (or *TTS* in this paper) has to be chosen from the critical range of the density in MFDs depicted in Figure 7 (a). From Figure 7 (a), a critical average density of 44 veh/km is derived, leading to $\hat{TTS} = 12.44(\text{km}) \times 44(\text{veh}/\text{km}) \approx 550$ veh introduced as a set-value to the PI feedback regulator (14). As in (Keyvan-Ekbatani et al., 2015a) $K_P = 20$ h^{-1} and $K_I = 5$ h^{-1} has been chosen as the controller gain values. As shown in Figure 7 (d), the controller successfully maintained the density or *TTS* close to the $\hat{TTS}$ and consequently kept the network performing near its capacity level. The average network flow enhancement under perimeter control strategy is much more obviously recognized by comparing the black curve with the grey continuous line in Figure 6.

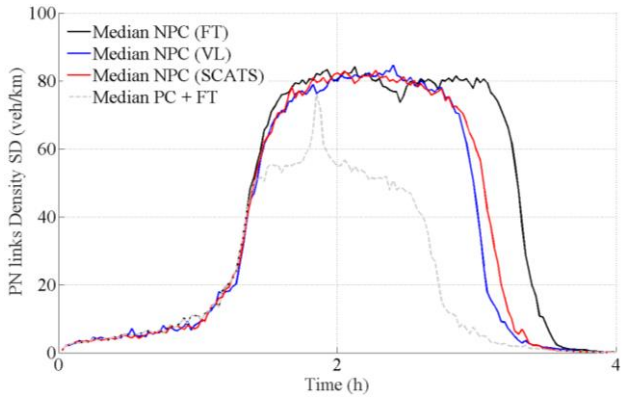

Figure 9 PN homogeneity in NPC (FT, VL and SCATS) and (PC + FT) scenarios

The MFD for the PN utilizing the MFD-based control strategy (PC + FT) is illustrated in Figure 8 (d). The figure shows a remarkable improvement in the link density SD (primarily experiencing the green and the yellow regions of SD as opposed to red regions) compared to NPC (FT) during the peak period and consequently a more homogenous spread of congestion over the network. Implementing the perimeter control strategy stopped the increase of traffic density: the controller (14) showed an excellent performance in keeping the network density within the range of the critical density (40 – 45 veh/km) and consequently maximized throughput in the range of 370-400 veh/h during the peak period for loading and loading. Hence, the *nucleation effect* (Knoop et al, 2013) with severe congestion and attraction of more congestion was avoided by PC strategy. As a result, the hysteresis in the MFD for the PC + FT compared to the NPC (FT) has been reduced considerably. Figure 10 (a) provides a better illustration of the link density SD for PC + FT and NPC (FT) scenarios. The PN link density SD for the 10 different runs are indicated by the yellow (NPC scenarios) and blue (PC scenarios) dots. The red and dark blue lines represent the median values for the NPC and PC scenarios, respectively. A clear reduction in the link density SD occurred in the PC + FT scenario compared to the NPC (FT) (compare the blue and the red lines at $t \in [1.3, 3.7]$ h in Figure 10 (a)).

6.5. Traffic-responsive versus perimeter control

Figure 9 displays the density SD of the links within the PN for the NPC and (PC + FT) simulation scenarios. The continuous blue and red lines represent the median fit over the corresponding 10-run simulation data for the NPC (VL and SCATS) scenarios, respectively. Clear decrease in the density SD has been obtained in the (PC + FT) simulation scenario (shown by the grey dashed-line). The MFD-based perimeter control strategy succeeds in reducing the heterogeneous spread of congestion remarkably across the PN network compared to the traffic-responsive strategies. The substantial difference between the traffic-responsive control strategies and perimeter control is more visible at the onset of congestion (at around $t = 1.3$ h). This confirms the findings in the previous literature on inefficiency of traffic-responsive control strategies in distributing the density homogenously and consequently mitigating the congestion in the network under over-saturated traffic regime (Kouvelas et al., 2017).

For a better comparison between the performance of traffic-responsive and perimeter control strategies, the replications with the best network performance in NPC (SCATS) and the worst (PC + FT), among the 10 simulation runs, have been chosen. Figure 11 (a) and (b) compare the average network flow and PN link density SD time series, respectively, for the two aforementioned simulation scenarios. As can be observed in Figure 11 (a), in under-saturated traffic condition $t \in [0, 1.5]$ h, the NPC (SCATS) and PC + FT scenarios (shown by blue continuous line and red dashed-line, respectively) exhibit similar performance. Figure 11 (a) reveals that even the worst perimeter control (PC + FT) simulation run outperforms the best NPC (SCATS) in oversaturated traffic conditions (the red dashed-line maintains at a higher level during the peak period $t \in [1.7, 2.6]$ h). Consequently, the congestion in the worst PC + FT scenario dissolves much quicker than in the best traffic-responsive-controlled scenario NPC (SCATS). In addition, the PN link density SD has been remarkably reduced in the PC + FT scenario during the rush hour $t \in [1.7, 2.6]$ h, as displayed in Figure 11 (b).

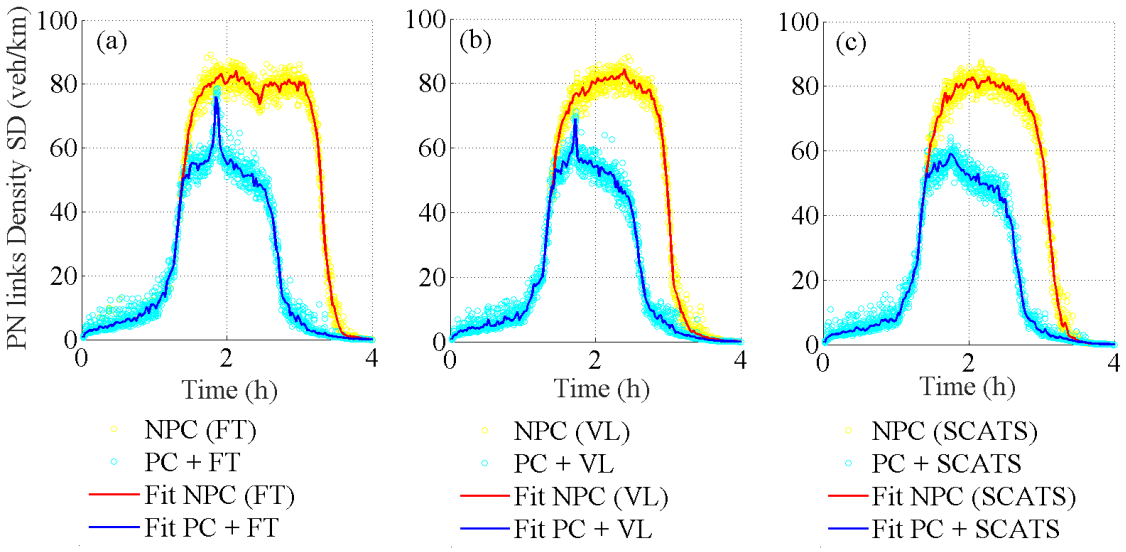

Figure 10 PN link density SD vs. time for (a) NPC (FT) and PC + FT; (b) NPC (VL) and PC + VL; (c) NPC (SCATS) and PC + SCATS

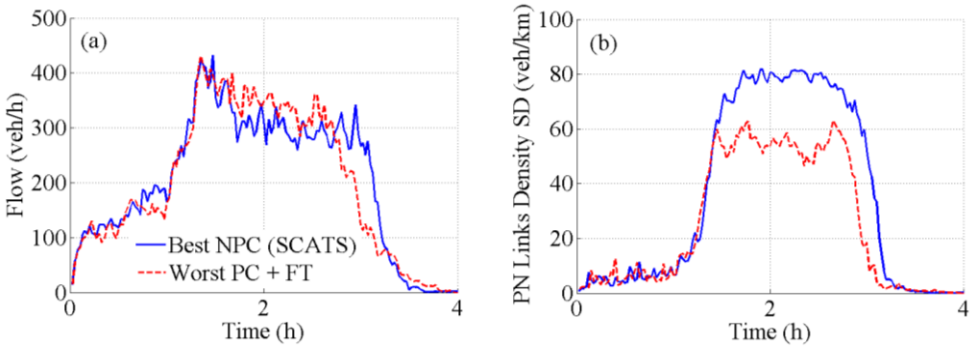

Figure 11 (a) PN average flow vs. time for the best NPC (SCATS) and the worst PC + FT performance among 10 runs; (b) PN link density SD vs. time for the best NPC (SCATS) and the worst PC + FT performance among 10 runs

6.6. Hierarchical traffic control versus PC + FT and NPC scenarios

This section explores the performance of the hierarchical traffic control strategy proposed in Figure 1 with scenarios in which only perimeter flow control or only traffic responsive signal control are implemented. Based on the MFDs plotted in Figure 7 (b) and (c), a critical density of 48 veh/km or $\hat{TTS} = 12.44(\text{km}) \times 48(\text{veh}/\text{km}) \approx 600$ has been considered as the controller set-value for both PC + VL and PC + SCATS. To the best of the authors' knowledge, this is the first time that the perimeter control concept has been applied in conjunction with traffic-responsive control strategies in order to control the boundary inflow into the subnetwork (at a *higher level*) and distribute the flow within the subnetwork (at a *lower level*). Utilizing the traffic-responsive control strategies (VL and SCTAS) combined with the perimeter control (PC) led to an obvious enhancement of the network capacity compared to the perimeter-control-only (PC + FT) scenario. This can be observed in Figure 8 (e) and (f). As illustrated in Figure 7 (e) and (f), the average network density has been kept around the critical density in both PC + VL and PC + SCATS scenarios. The scattered blue (+) inside the dashed red circle in Figure 7 (d-f) confirm that applying perimeter control with all three control strategies (FT, VL and SCATS) lead to higher network throughput during the whole 2 hours of loading period (the concentration of (+) are near critical capacity instead of high density values as in Figure 7 (a-c)). Note that the scattered (+) show the simulation runs for 10 different

replications. A deeper insight into the link density SD reveals that the PN in the PC + SCATS scenario experienced the lowest link density variability during the whole simulation period (see Figure 10).

Figure 12 displays the network average flow time series for all the 6 simulation scenarios. The median of the 10 simulation runs for the PC and NPC scenarios has been shown by the blue and red lines, respectively, in Figure 12 (the dots show the 10 simulation runs for the corresponding scenarios). The comparison between the NPC and PC lines in Figure 12 (a-c) unveiled that applying the perimeter control with any of the three control concepts (FT, VL and SCATS) is extremely beneficial in increasing the network capacity. In addition, congestion dissolves much faster in all the PC scenarios. The two-level control strategy proposed in this paper (perimeter control combined with the traffic-responsive SCATS) results in higher capacity compared to the other two PC scenarios.

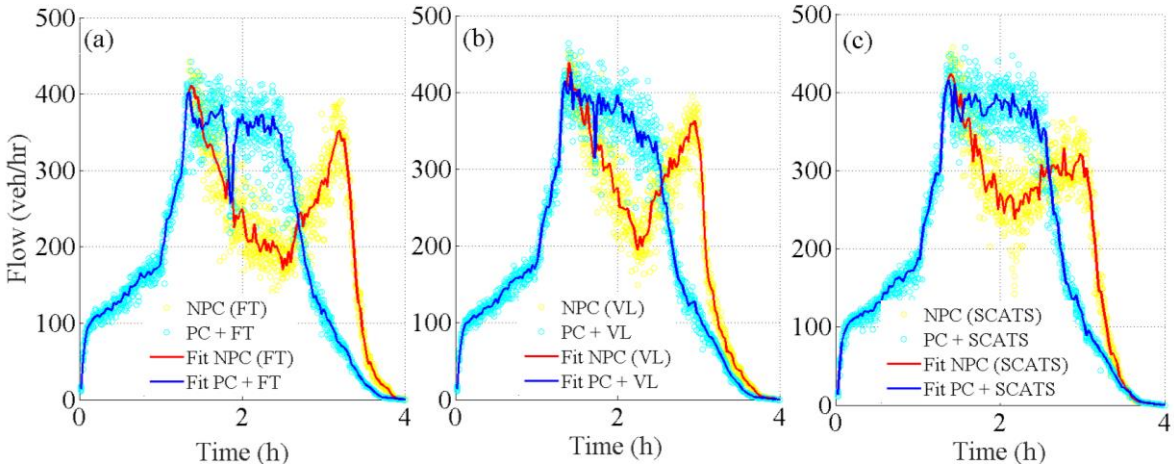

Figure 12 PN average flow vs. time for (a) NPC (FT) and PC + FT ;(b) NPC (VL) and PC + VL; (c) NPC (SCATS) and PC + SCATS

To assess the network performance at the global network level (i.e. the right figure in Figure 4), the average vehicle delay (s/km), average virtual queue (veh) and total time spent are considered. As addressed earlier in this paper, the NPC (FT) is selected as the base scenario for the comparative simulation study.

The boxplot displayed in Figure 13 summarizes the overall network delay values and average virtual queue (all replications) for the six scenarios (NPC (FT, VL and SCATS) and PC (FT, VL and SCATS)) 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 seen in Figure 13 (a), delay is remarkably reduced in the PC + SCATS and PC + VL scenarios compared to the NPC (FT), by an average of 51% and 52%, respectively. Moreover, the reduced height of the box in PC + SCATS and PC + VL scenarios indicates lower variability of the simulation results compared to the NPC (FT) scenario. Compared to NPC (FT), the SD of average delay reduced by 70% and 71% in PC + VL and PC +SCATS scenarios, respectively.

Figure 13 (b) shows that implementing perimeter control leads to higher virtual queues in general. However, interestingly, due to higher network critical density (see Figure 6 (b)) in the hierarchal traffic control scenario, fewer cars are forced to queue on the perimeter of the PN. Consequently, the length of the average virtual queue in PC + VL and PC + SCATS is reduced remarkably compared to the PC + FT scenario. This is indeed an important achievement, since the spillback (gridlock) probability at the gated links might be considerably decreased by reducing the queue length at the border of protected area.

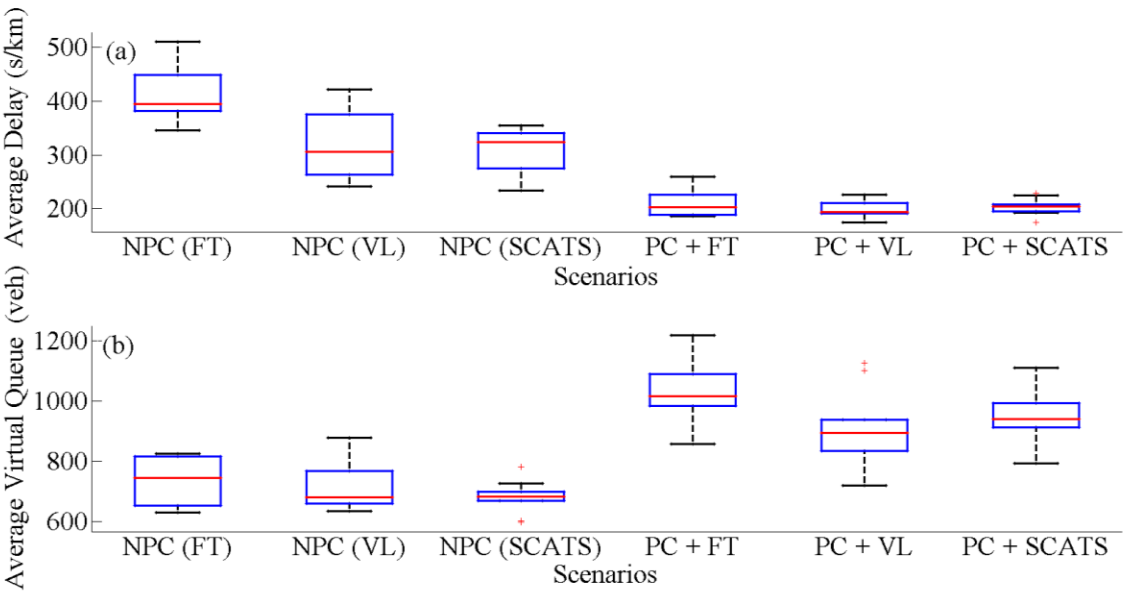

Figure 13 Boxplot of the overall network performance for NPC and PC scenarios; (a) average delay (s/km); (b) average virtual queue (veh)

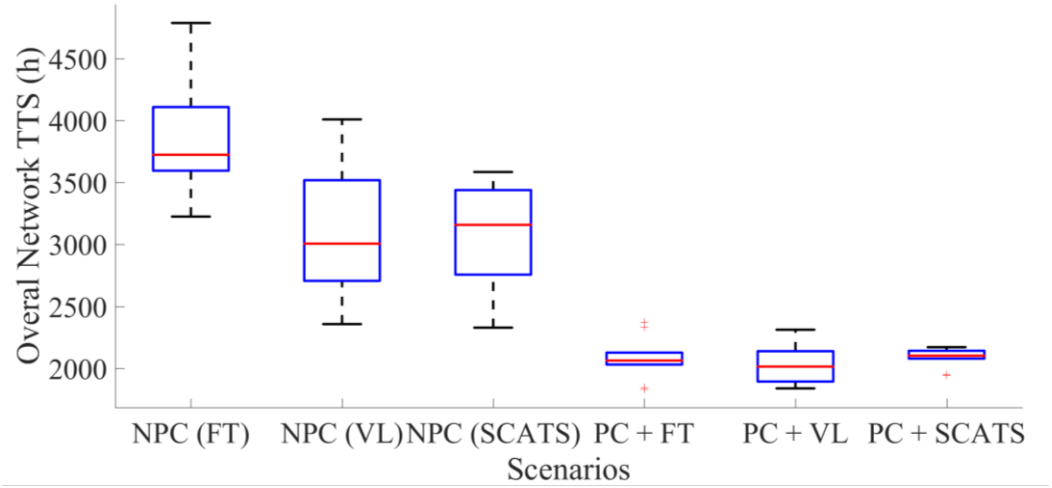

Figure 14 Boxplot of the overall network Total Time Spent (*TTS*) for NPC and PC scenarios

Figure 14 displays the *TTS* in (h) for the entire network at the end of 4-h simulation for each scenario over the 10 different simulation runs. As expected, the median of TTS values reduced significantly after applying perimeter control to the NPC scenarios. The height of the box in PC + SCATS scenario reduced remarkably compared to the other PC scenarios which shows lower variability of the simulation results. Note that with similar improvement of the overall *TTS*, both hybrid real-time control strategies (PC+VL and PC+SCATS) reduced queuing at the boundary of the entire network (see Figure 13) and consequently the overall delay.

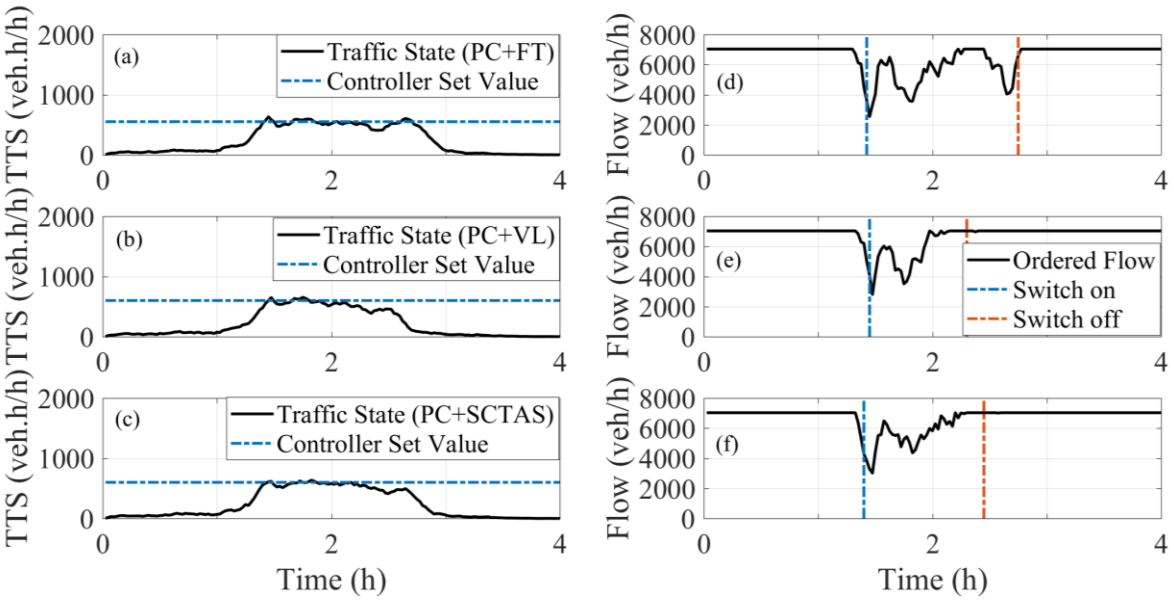

Figure 15 Controller action for a simulation run in three different perimeter control scenarios (PC+FT (a and d), PC+VL (b and e) and PC+SCATS (c and f)): left column (a-c) real-time traffic state (i.e. TTS) over the 4-h simulation; right column (d-f) controller ordered flow during gating

The controller action in PC scenarios for a single replication have been depicted in Figure 15. Focusing on the left column (a-c) in Figure 15, the PI controller (10) has successfully kept the real-time traffic state (i.e. TTS) within the protected network close to the set values in all PC scenarios. Figure 15 (d-f) illustrates the ordered flow from (10) in PC+FT, PC+VL and PC+SCATS scenarios. Once the *TTS* overshoot occurs (e.g. at 1.4h in Figure 15 (a)) the controller orders lower flow to reduce the entering flow to PN. Utilizing the traffic-responsive control strategies (VL and SCATS), the gating period (indicated by the vertical dashed lines) reduced compared to the scenario PC+FT. This might be another important consequence of more homogenous traffic conditions applying traffic responsive control strategies within the PN (as seen in Figure 9).

7. Discussion and Conclusions

This paper considers the joint implementation of two unique urban traffic control strategies: locally adaptive traffic signal control and perimeter gating of a protected network. The former modifies timings at individual intersections in response to real-time traffic fluctuations. In general, the goal of adaptive traffic signal control is to provide more green time to the approaches that need it most, which usually are those that are the most congested. The latter strategy limits vehicle entries into a protected network to minimize the probability that the network becomes congested. It leverages the existence of well-defined MFDs that relate network production with current traffic state. By limiting vehicle entries, the network is allowed to to operate with the maximum vehicle throughput possible, which is represented by the capacity state of the MFD.

Previous research has found that each of these strategies (individually) can improve overall traffic conditions within a network. However, adaptive traffic signal control cannot prevent the formation of congestion and struggles to perform well in congested networks. To show this, lower and higher demand level i.e. 80% and 110% of the OD matrix, respectively have been introduced for the different simulation scenarios and the network performance have been studies under the corresponding demand level.

7.1. PN traffic condition under low and high demand levels for NPC and PC scenarios

Figure 16 displays the MFDs of the PN for the 80% demand profile. As already expected from the literature, the traffic-responsive control strategies are outperforming the fixed-time (FT) control strategy in under-saturated/capacity traffic conditions. The two traffic-responsive strategies (i.e. VL and SCATS) provided higher maximum flow compared to the FT strategy in the PN and kept the network average density close to the critical average density. This means, for this demand level, perimeter control is not required since the VL and SCATS are performing efficiently.

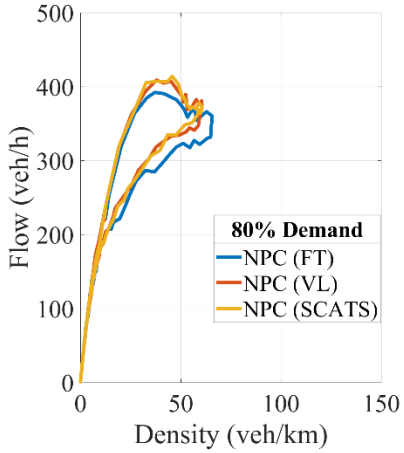

Figure 16 MFD of the PN for lower demand level (80% of OD demand profile) for the NPC scenarios (averaged over 10 replications for each scenario)

Figure 17 illustrates the MFDs for the oversaturated traffic conditions (i.e. 110% demand profile). As seen in Figure 17 (left) none of the NPC strategies could prevent the PN from oversaturation (large average densities and drop in the average network flow). Applications of PC could maintain the network average density close to the critical density range and keep the network flow at the maximum level. As shown in the Figure 17 (right) even in very highly saturated traffic conditions the robust structure of the PI feedback regulator could efficiently perform and maximize the network throughput.

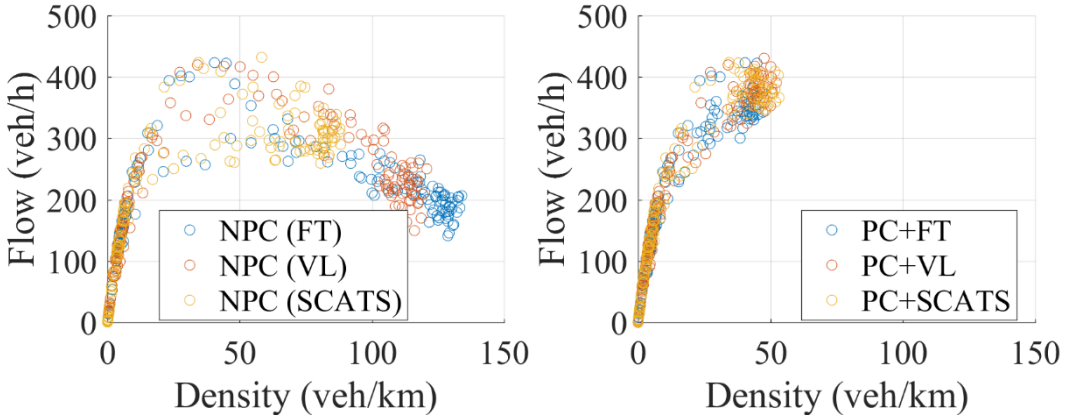

Figure 17 MFD of the PN for a high demand level (110% of OD demand profile) for the NPC scenarios (left); and PC (right)

7.2. Benefits of combining PC with traffic-responsive strategies at the local level (intersections within PN)

The simulation study in this paper and the historical data show that normally the main arterial (indicated by the bold arrow in Figure 18) in protected network of the city of Chania in Greece turns into oversaturated traffic condition during the peak period at the initial stages. Later, the congestion propagates to other intersections in the surrounding area (shown by the red circle) and creates partial gridlocks which deteriorate the traffic condition in the network (entering the right hand side of the MFD).

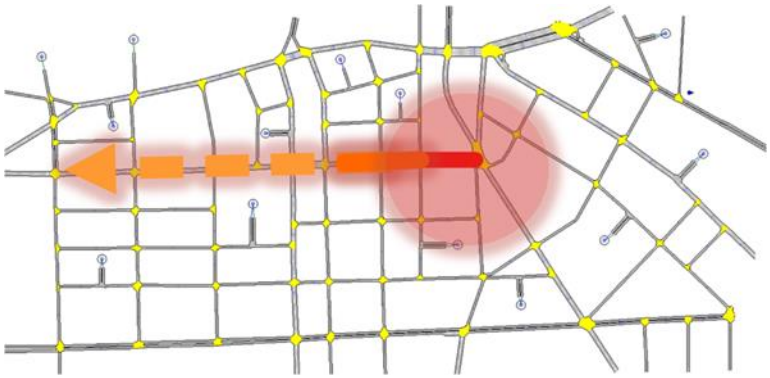

Figure 18 Critical intersection and congestion propagation during peak period within the PN

To investigate more closely the traffic condition at the local level, an intersections within the PN has been chosen which has normally the highest flow rate and largest congestion potential at the early stage during the peak period (the arterial link leading to the aforementioned intersection has been indicated by red bold line on the Figure 18. The occupancy values of the red-indicated link has been illustrated in Figure 19 for the six different simulation scenarios for one replication.

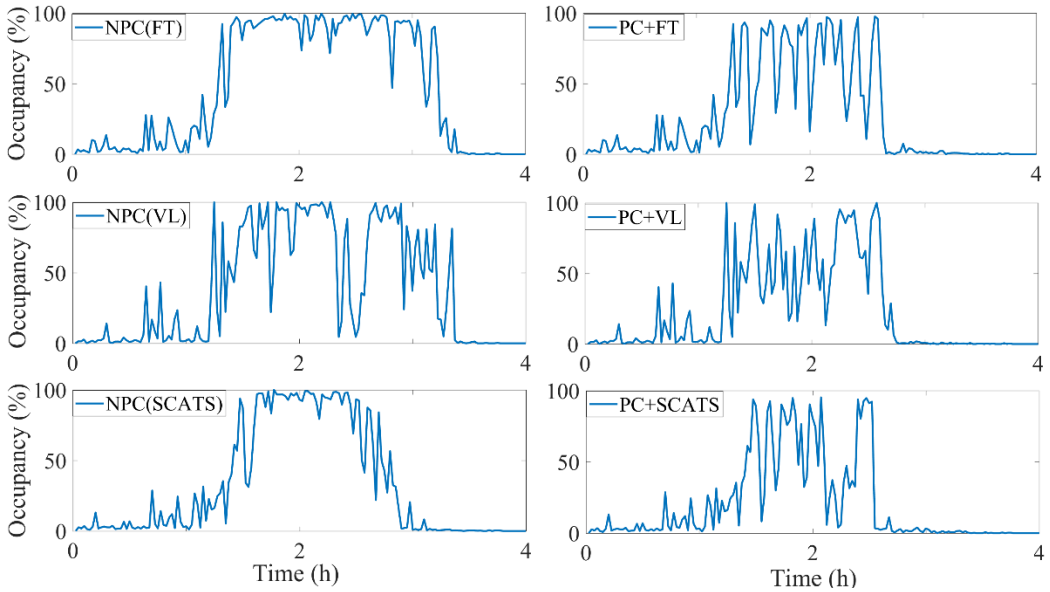

Figure 19 Occupancy values for NPC scenarios (left column); occupancy of the red-indicated link in PN for PC scenarios (right column)

As shown in Figure 19, for the three NPC scenarios, the occupancy values of the red-indicated link on Figure 18 are mostly at the highest level (i.e. 100%) during the peak period. In addition, the duration of congestion is around 2 hours (1.5h to 3.5h) for the NPC (FT) and the NPC (VL) scenarios. The congestion dissolves earlier in NPC (SCATS).

The occupancy values for the link has been reduced significantly in all PC scenarios (right column of Figure 19). More interestingly, the combination of perimeter control and the traffic-responsive strategies resulted in more efficient traffic management at the critical link within the PN by not only reducing the occupancy values (i.e. the congestion level) but also shortening the congestion duration (congestion dissolves at around 2.5h in all PC scenarios). This confirms the benefit of the combined strategy not only at the network but also at the link level.

7.3. Further discussions

Here, we have shown that the combination of these two strategies can be even more beneficial than either of these strategies operating in isolation. Specifically, the perimeter flow control allows a network to operate in uncongested states for which the locally adaptive traffic signals are most beneficial. The result is networks that operate with more homogeneous distributions of congestion and at higher efficiencies. This results in shorter peak periods and faster dissipation of congestion. More interestingly, adaptive signal control actually improves a network's capacity and the range of densities for which it is observed. This allows the perimeter flow control to be less restrictive, which results in lower overall travel delays and less queuing on the periphery of the protected network. This latter effect is particular useful in multi-region networks for which queues on the boundary of the protected region might create congestion problems in other regions. Thus, the combination of locally adaptive signals and perimeter flow control is found to lower network delays, reduce boundary queues (on average), promote higher travel speeds speeds and increase overall vehicle throughputs.

The results of this study were obtained using micro-simulations of realistic traffic networks. A limitation of this work is that only two illustrative adaptive signal control schemes are considered. These two illustrative schemes were selected because they mimic the basic goal of most adaptive systems: provide green time to the more congested approaches that need more throughput. Numerous advanced adaptive traffic signal systems have been either proposed (e.g. Yu et al. (2018); Zheng et al. (2018)) or are in implementation throughout the world. Thus, while this study finds that perimeter flow control generally outperforms adaptive signal control, these results might not hold for all adaptive signal systems. It is possible that some adaptive signal control schemes perform much better when a network is congested, particularly if the strategy specifically considers queue spillbacks and oversaturation within the network. Nevertheless, we think combination of better performing adaptive signal systems with perimeter flow control is very likely to have even more benefits than those illustrated here. At the end, we would like to conclude that the existing urban networks operating with traffic-responsive control strategies (e.g. SCATS, TUC, etc.) will benefit if they implement gating at a higher level of traffic control. Future research direction will be comparing the PC strategy with more advanced adaptive traffic control strategies like TUC in a multi-region network. The novel real-time deep-learning-based queue estimation approach (Lee et al., 2018) will be integrated to the PC framework. The convincing outcome of this simulation study motivates the real-field implementation of gating/perimeter control in the cities, without even changing the existing adaptive control strategies.

References

Aboudolas, K., Papageorgiou, M., Kosmatopoulos, E., 2009. Store-and-forward based methods for the signal control problem in large-scale congested urban road networks. *Transportation Research Part C* 17(2), 163-174.

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.

Correia, G., Viegas, J.M., 2011. Carpooling and carpool clubs: Clarifying concepts and assessing value enhancement possibilities through a stated preference web survey in Lisbon, Portugal. *Transportation Research Part A: Policy and Practice* 45, 81–90.

Daganzo, C.F., 2007. Urban gridlock: Macroscopic modeling and mitigation approaches. *Transportation Research Part B* 41(1), 49-62.

Diakaki, C., Papageorgiou, M., Aboudolas, K., 2002. A multivariable regulator approach to traffic-responsive network-wide signal control. *Control Engineering Practice* 10(2), 183-195.

Drake, J. S., Schöfer, J. L., May, A., 1967. A Statistical analysis of apeed density hypotheses. In Proceedings of the Third International Symposium on the Theory of Traffic Flow (L. C. Edie, R. Herman, and R. Rothery, eds.), Elsevier North-Holland, New York.

Du, J., Rakha, H., Gayah, V., 2015. Deriving macroscopic fundamental diagrams from probe data: Issues and proposed solutions. Transportation Research part C: Emerging Technologies 66, 136-149.

Edie, L.C., 1965. Discussion of traffic stream measurements and definitions, *the 2nd International Symposium on the Theory of Traffic Flow*.

Farges, J.-L., Henry, J.-J., Tuffal, J., 1983. The PRODYN real time traffic algorithm, *4th IFAC Symposium on Transportation Systems*, 307-312.

Gartner, N.H., 1983. *OPAC: A demand-responsive strategy for traffic signal control*.

Gayah, V., Daganzo, C., 2011a. Effects of turning maneuvers and route choice on a simple network. *Transportation Research Record* 2249, 15-19.

Gayah, V., Dixit, V., 2013. Using mobile probe data and the macroscopic fundamental dagram to esimate network densities. *Transportation Research Record* 2390, 76-86.

Gayah, V.V., Daganzo, C.F., 2011b. Clockwise hysteresis loops in the macroscopic fundamental diagram: An effect of network instability. *Transportation Research Part B* 45(4), 643-655.

Gayah, V.V., Gao, X., Nagle, A.S., 2014. On the impacts of locally adaptive signal control on urban network stability and the Macroscopic Fundamental Diagram. *Transportation Research Part B* 70(0), 255-268.

Geroliminis, N., Zheng, N., Ampountolas, K., 2014. A three-dimensional fundamental diagram for mixed bi-modal urban networks. *Transportation Research Part C: Emerging Technologies* 42, 168–181.

Geroliminis, N., Haddad, J., Ramezani, M., 2013. Optimal perimeter control for two urban regions with macroscopic fundamental diagrams: A model predictive approach. *IEEE Transactions on Intelligent Transportation Systems* 14, 348-359.

Geroliminis, N., Daganzo, C.F., 2008. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B* 42(9), 759-770.

Girault, J. T., Gayah, V. V., Guler, I., Menendez, M., 2016. Exploratory analysis of signal coordination impacts on macroscopic fundamental diagram. *Transportation Research Record: Journal of the Transportation Research Board, 2560*, 36-46.

Godfrey, J., 1969. The mechanism of a road network. *Traffic Engineering and Control 11, 323–327.* 11, 323–327.

Haddad, J., Ramezani, M., Geroliminis, N., 2013. Cooperative traffic control of a mixed network with two urban regions and a freeway. *Transportation Research B* 54, 17-36.

Haddad, J., Shraiber, A., 2014. Robust perimeter control design for an urban region. *Transportation Research Part B* 68, 315-332.

Haddad, J. and Zheng, Z., 2018. Adaptive perimeter control for multi-region accumulation-based models with state delays. Transportation Research Part B: Methodological.

Han, K., Liu, H., Gayah, V.V., Friesz, T.L., Yao, T., 2016. A robust optimization approach for dynamic traffic signal control with emission considerations. Transportation Research Part C: Emerging Technologies 70, 3-26.

Hunt, P., Robertson, D., Bretherton, R., Royle, M., 1982. The SCOOT on-line traffic signal optimization technique. *Traffic Engineering and Control* 23, 190-192.

Keyvan-Ekbatani, M., Gao, X., Gayah, V.V. and Knoop, V.L., 2016a. Combination of traffic-responsive and gating control in urban networks: Effective interactions. *In Proceedings of the 95th annual meeting of the Transportation Research Board*, Washington (USA).

Keyvan-Ekbatani, M., Carlson, R.C., Knoop, V.L., Hoogendoorn, S.P. and Papageorgiou, M., 2016b, November. Queuing under perimeter control: Analysis and control strategy. In Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference, 1502-1507.

Keyvan-Ekbatani, M., Yildirimoglu, M., Geroliminis, N., Papageorgiou, M., 2015a. Multiple concentric gating traffic control in large-scale urban networks. *IEEE Transactions on Intelligent Transportation Systems*, *16*(4), 2141-2154.

Keyvan-Ekbatani, M., Papageorgiou, M., Knoop, V.L., 2015b. Controller design for gating traffic control in presence of time-delay in urban road networks. *Transportation Research Part C: Emerging Technologies* 59, 308-322.

Keyvan-Ekbatani, M., Papageorgiou, M., Papamichail, I., 2014. Perimeter traffic control via remote feedback gating. *Procedia – Social and Behavioral Sciences* 111, 645-653.

Keyvan-Ekbatani, M., Papageorgiou, M., Papamichail, I., 2013. Urban congestion gating control based on reduced operational network fundamental diagrams. *Transportation Research Part C* 33, 74-87.

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* 46(10), 1393-1403.

Knoop, V., Van Lint, J., Hoogendoorn, S., 2015. Traffic dynamics: Its impacts on the Macroscopic Fundamental Diagram. *Physica A* 438, 236-250.

Knoop V.L., J.W.C. van Lint, J. Vries, L. Kester, I. Passchier, 2013, Relationship between application scale and maximum time latency in intelligent transport solutions, Transportation Research Records No. 2380, 1-9.

Kouvelas, A., Saeedmanesh, M. Geroliminis, N., 2017. Enhancing model-based feedback perimeter control with data-driven online adaptive optimization. *Transportation Research Part B* 96, 26-45.

Leclercq, L., Chiabaut, N., Trinquier, B., 2014. Macroscopic fundamental diagrams: A cross-comparison of estimation methods. *Transportation Research Part B* 62(1), 1-12.

Leclercq, L., Geroliminis, N., 2013. Estimating MFDs in simple networks with route choice. *Transportation Research Part B* 57, 468-484.

Leclercq, L., Parzani, C., Knoop, V., Amourette, J., Hoogendoorn, S., 2015. Macroscopic traffic dynamics with heterogeneous route patterns. *Transportation Research Part C: Emerging Technologies* 59, 292-307.

Lee, S., Xie, K., Ngoduy, D., Keyvan-Ekbatani, M., Yang, H., 2018. Real-time lane-based queue length prediction: a deep learning approach. *ISTS & IWTDCS 2018*.

Lo, H.K., 1999a. A cell-based traffic control formulation: Strategies and benefits of dynamic timing plans. *Transportation Science* 35(2), 148-164.

Lo, H.K., 1999b. A novel traffic signal control formulation. *Transportation Research Part A* 33(6), 433-448.

Lowrie, P.R., 1982. SCATS: the Sydney coordinated adaptive traffic system—principles, methodology, algorithms, *IEEE International Conference on Road Traffic Signaling*, 67-70.

Mahmassani, H., Williams, J., Herman, R., 1987. Performance of urban traffic networks *The 10th International Symposium on Transportation and Traffic Theory*, 1-20.

Mahmassani, H., Williams, J.C., Herman, R., 1984. Investigation of network-level traffic flow relationships: some simulation results. *Transportation Research Record* 971, 121-130.

Mahmassani, H.S., Saberi, M., Zockaie, A., 2013. Urban network gridlock: Theory, characteristics, and dynamics. *Transportation Research Part C* 36(0), 480-497.

Mehr, N., Lioris, J., Horowitz, R. and Pedarsani, R., 2017. Joint perimeter and signal control of urban traffic via network utility maximization. In *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)* (pp. 1-6).

Nagle, A., Gayah, V., 2014. The accuracy of network-wide traffic state esimations using mobile probe data. *Transportation Research Record* 2421, 1-11.

Olszewski, P., Fan, H., Tan, Y., 1995. Area-wide traffic speed-flow model for the Singapore CBD. *Transportation Research Part A* 29(4), 273-281.

Ramezani, M., Nourinejad, M., 2017. Dynamic modeling and control of taxi services in large-scale urban networks: A macroscopic approach. *Transportation Research Procedia 23*,41-60.

Ramezani, M., Haddad, J., Geroliminis, N., 2015. Dynamics of heterogeneity in urban networks: aggregated traffic modeling and hierarchical control. *Transportation Research Part B* 74, 1-19.

Saberi, M., Mahmassani, H.S., Zockaie, A., 2014. Network capacity, traffic instability, and adaptive driving: findings from simulated urban network experiments. *EURO Journal on Transportation and Logistics* 3, 289-308.

Schrank, D., Eisele, B., Lomax, T., 2012. TTI's 2012 urban mobility report. *Texas A&M Transportation Institute. The Texas A&M University System*.

Seborg, D., Edgar, T.F., Mellichamp D.A., 1989. Process Dynamics and Control. New York: Wiley.

Slavin, C., Feng, W., Figliozzi, M., Koonce, P., 2013. A statistical study of the impacts of SCATS adaptive traffic signal control on traffic and transit performance. *Transportation Research Record* 2356, 117-126.

Smeed, R., 1966. Road capacity of city centers. *Traffic Engineering and Control* 9, 455-458.

TSS Transport Simulation Systems, 2016. AIMSUN User Manual Version 8, Barcelona, Spain.

Tsubota, T., Bhaskar, A., Chung, E., 2014. Macroscopic fundamental diagram for Brisbane, Australia: Empirical findings on network partitioning and incident detection. *Transportation Research Record* 2421, 12-21.

Ukkusuri, S., Doan, K., Aziz, H.M.A., 2013. A bi-level formulation for the combined dynamic equilibrium based traffic signal control. *Procedia-Social and Behavioral Sciences* 80, 729-752.

Van de Weg, G.S., Keyvan-Ekbatani, M., Hegyi, A., Hoogendoorn, S.P., 2016. Urban Network Throughput Optimization via Model Predictive Control Using the Link Transmission Model. *Transportation Research Board 95th Annual Meeting* (No. 16-1512).

Verhoef, E., 2002. Second-best congestion pricing in general networks: heuristic algorithms for finding second-best optimal toll levels and toll points. *Transport Research Part B* 36 (8), 707–729.

Vigos, G., Papageorgiou, M., 2010. A simplified estimation scheme for the number of vehicles in signalized links. *IEEE Transactions on Intelligent Transportation Systems 11*, 312-321.

Yang, K., Zheng, N. and Menendez, M., 2017. Multi-scale perimeter control approach in a connected-vehicle environment. *Transportation Research Part C (ISTTT22 special issue)*.

Yu, C., Feng, Y., Liu, H. X., Ma, W., & Yang, X. (2018). Integrated optimization of traffic signals and vehicle trajectories at isolated urban intersections. Transportation Research Part B: Methodological, 112, 89-112.

Yildirimoglu, M., Ramezani, M., Geroliminis, N., 2015. Equilibrium analysis and route guidance in large-scale networks with MFD dynamics. *Transportation Research Part C 59*, 404-420.

Zahavi, Y., 1972. Traffic performance evaluation of road networks by the a-relationships, Parts 1 and 2. *Traffic Engineering and Control* 14, 228–231, 292-293.

Zargari, S.A., Dehghani, N. and Mirzahossein, H., 2018. Optimal traffic lights control using meta heuristic algorithms in high priority congested networks. Transportation letters, 10(3), pp.172-184

Zhang, L., Garoni, T.M., de Gier, J., 2013. A comparative study of Macroscopic Fundamental Diagrams of arterial road networks governed by adaptive traffic signal systems. *Transportation Research Part B* 49(0), 1-23.

Zheng, J., Sun, W., Huang, S., Shen, S., Yu, C., Zhu, J., Liu, B., Liu, HX, 2018. Traffic Signal Optimization Using Crowdsourced Vehicle Trajectory Data. Transportation Research Board 97th Annual Meeting Transportation Research Board.

Zheng, N., Geroliminis, N., 2013. On the distribution of urban road space for multimodal congested networks. *Transportation Research B* 57, 326-341.