FEEDBACK PERIMETER CONTROL FOR MULTI-REGION AND HETEROGENEOUS CONGESTED CITIES

Konstantinos Aboudolas
École Polytechnique Fédérale de Lausanne (EPFL)
School of Architecture, Civil & Environmental Engineering
Urban Transport Systems Laboratory
GC C2 390 (Bâtiment GC), Station 18
1015 Lausanne, Switzerland
Phone: +41-21-69-32964
E-mail: konstantinos.ampountolas@epfl.ch

Nikolas Geroliminis
École Polytechnique Fédérale de Lausanne (EPFL)
School of Architecture, Civil & Environmental Engineering
Urban Transport Systems Laboratory
GC C2 389 (Bâtiment GC), Station 18
1015 Lausanne, Switzerland
Phone: +41-21-69-32481
E-mail: nikolas.geroliminis@epfl.ch

*Corresponding Author
Number of words: 6,300 + 1,250 (5 Figures) + 250 (1 Table) = 7,800

ABSTRACT

It was recently observed from empirical data that by aggregating the highly scattered plots of flow versus density from individual loop detectors for city regions with homogeneous spatial distribution of congestion, the scatter almost diminishes and a well-defined Macroscopic Fundamental Diagram (MFD) exists between space-mean flow and density. This result can be of great importance to unveil simple perimeter control policies in such a way that maximizes the network capacity and outflow. Single region perimeter control might be suboptimal if there is a significant number of destinations outside the region of analysis or if the city is heterogeneously congested. This paper integrates an MFD modeling to perimeter control optimization for large-scale cities with multiple centers of congestion, if these cities can be partitioned in a small number of homogeneous regions. Perimeter control actions may be computed in real-time through a linear multivariable feedback regulator or a linear multivariable integral feedback regulator. The impact of the perimeter control actions to a three-region real urban network (Downtown of San Francisco) is demonstrated via micro-simulation. A key advantage of this approach is that it does not require high computational effort and future demand data if the state of each region can be observed.

Keywords: Macroscopic Fundamental Diagram, Perimeter Control, Multivariable Feedback Regulators, Multi-region Cities
INTRODUCTION

Analysis of traffic theory and modeling of vehicular congestion have similarities with other physical systems like fluid mechanics, many particles systems and the like, which inspired many scientists to create analogies and fundamental laws inspired from physics. One of the most broadly used traffic laws is the *Fundamental Diagram* (FD) which was initially observed for a stretch of highway (1) provides a steady-state relationship between traffic variables (speed, density, and flow). Although quite capable in providing a coarse description of main traffic features (e.g., formation and dissolution of shockwaves), the FD is inadequate in describing some complex traffic patterns such as stop-and-go waves, capacity drop phenomena, traffic oscillations, which create non-steady states below the upper bound of an FD, and it also contains significant experimental errors, especially in the congested regime, see e.g. (2, 3).

Recently, however, it was observed from empirical data (4) that by aggregating the highly scattered plots of flow versus density from individual loop detectors, the scatter almost disappeared and a well-defined FD exists between space-mean flow and density. This field experiment in downtown Yokohama, Japan, showed that if congestion is homogeneously distributed (i) urban neighborhoods approximately exhibit a *Macroscopic Fundamental Diagram* (MFD) relating the number of vehicles to space-mean speed (or flow), (ii) there is a robust linear relation between the neighborhood’s average flow and its total outflow (rate vehicles reach their destinations) and (iii) the MFD is a property of the network infrastructure and control and not of the demand. The first proposition of a macroscopic relationship between average network flow and density with an optimum accumulation belongs to Godfrey (5), while other empirical and micro-simulation studies for MFDs can be found in (6, 7, 8, 9, 10). Despite the recent findings for well-defined MFDs with low scatter in homogeneously congested networks, in reality many urban transportation networks are heterogeneous with different congestion conditions and these curves should not be a universal law. For example, recent simulation and empirical findings (3, 11, 12) have shown that strong hysteresis phenomena exist in MFDs for urban and freeway networks.

These results can be of great importance to unveil simple perimeter control policies in such a way that maximizes the network throughput/capacity and/or the density range of the capacity. The general idea of a perimeter control policy is to “meter” the input flow to the system and to hold vehicles outside the system if necessary (13). Recently, this general idea has been an issue of investigations for single-region (14) and two-region (15) systems. Single region perimeter control (13, 14) might be suboptimal if there is a significant number of destinations outside the region of analysis. It might not be equitable as well given that all the penalties of waiting are transferred to external vehicles moving from outside to the study region. On the other hand, the authors in (15) developed a Model Predictive perimeter Control (MPC) strategy for a hypothetical closed-loop network with two regions. However, MPC requires that future demands be predicted, and is thus impractical for on-line use. In fact, if prediction is not successful (e.g. because of demand uncertainty) optimal solutions obtained with MPC can result in performance that is even worse than without prediction.

One can model a city as a single or multi-reservoir system (see Figures 1(a) and 1(b), respectively) depending on the geometry, the spatial distribution of congestion and the trip destinations among the city. The requirement for homogeneity in traffic loads and slow variations in time should determine the number of required reservoirs and the time scale of the model. Networks with an uneven and inconsistent distribution of congestion may exhibit traffic states that

TRB 2013 Annual Meeting
FIGURE 1 A city modeled as (a) single-reservoir system and (b) multi-reservoir system.

are much too scattered to line along an MFD and require partitioning (16). Moreover, a simple perimeter control policy may induce uneven distribution of vehicles in the reservoirs, and, as a consequence, may invalidate the homogeneity assumption of traffic loads within the reservoirs and degrade the total network throughput and efficiency. In this work we put some effort to deal with these important issues of efficiency, heterogeneity and equity.

To address the aforementioned issues, this paper develops the dynamics of cities with more complicated structure. More specifically, a generic mathematical model of an $N$-reservoir city with well-defined MFDs for each reservoir is presented first. Two modeling variations lead to two alternative optimal control methodologies for the design of perimeter control strategies that aim at distributing the accumulation in each reservoir as homogeneously as possible, and maintaining the rate of vehicles that are allowed to enter each reservoir around a desired point, while the system’s throughput is maximized. Based on the two control methodologies, perimeter control actions may be computed in real-time through a linear multivariable feedback regulator or a linear multivariable integral feedback regulator, respectively. To this end, the heterogeneous network of the Downtown of San Francisco is partitioned into three homogeneous regions that approximately exhibit well-defined MFDs. These MFDs are then used to design the linear multivariable feedback regulator. Finally, the impact of the perimeter control actions to the three reservoirs and the whole network is demonstrated via simulation by the use of the corresponding MFDs.

DYNAMICS FOR A CITY PARTITIONED IN $N$ RESERVOIRS

Consider a city partitioned in $N$ reservoirs (Figure 1(b)). Denote by $i = 1, \ldots, N$ a reservoir in the system, and let $n_i(t)$ be the accumulation of vehicles in reservoir $i$ at time $t$; $n_i,\text{max}$ be the maximum accumulation of vehicles in reservoir $i$. We assume that for each reservoir $i = 1, \ldots, N$ there exists an MFD, $O_i(n_i(t))$, between accumulation $n_i$ and output $O_i$ (number of trips exiting reservoir $i$ per unit time either because they finished their trip or because they move to another reservoir), which describes the behavior of the system when it evolves slowly with time $t$.

Let $q_{i,\text{in}}(t)$ and $q_{i,\text{out}}(t)$ be the inflow and outflow in reservoir $i$ at time $t$, respectively; $S_i$ be the set of origin reservoirs whose outflow will go to destination reservoir $i$ (including reservoir $i$ in case that reservoir $i$ is reachable from the perimeter). Also, let $d_i(t)$ be the uncontrolled
traffic demand (disturbances) in reservoir $i$ at time $t$. Note that $d_i(t)$ includes both internal (off-street parking for taxis and pockets for private vehicles) and external (non-controlled) inflows. The conservation equation for each reservoir $i = 1, \ldots, N$ reads:

$$\frac{dn_i(t)}{dt} = q_{i,\text{in}}(t) - q_{i,\text{out}}(t) + d_i(t).$$  

(1)

Since the system of each reservoir evolves slowly with time, we may assume that the outflow $q_{i,\text{out}}(t)$ is function of the output $O_i(n_i(t))$ (the MFD), where output $O_i(n_i(t))$ is the sum of the exit flows from reservoir $i$ to reservoir $j$, plus the internal output (internal trip completion rates at $i$). If $i$ and $j$ are two reservoirs sharing a common boundary, we denote by $\beta_{ji}$ ($j \neq i$) the rate of vehicles in reservoir $j$ that are allowed to enter reservoir $i$ and by $\beta_{ii}$ the rate of vehicles in the perimeter of the network allowed to enter reservoir $i$ (see Figure 1(b)). The inflow to reservoir $i$ is given by

$$q_{i,\text{in}}(t) = \sum_{j \in S_i} \beta_{ji}(t - \tau_{ji})O_j(n_j(t))$$

(2)

where $\beta_{ji}(t - \tau_{ji})$ are the input variables from reservoir $j$ to reservoir $i$ at time $t$, to be calculated by the perimeter controller, and $\tau_{ji}$ is the travel time needed for vehicles to approach reservoir $i$ from origin reservoir $j$. Without loss of generality, we assume that $\tau_{ji} = 0$, i.e., vehicles can immediately get access to the reservoirs of the network. This assumption can be readily removed by introducing additional auxiliary variables. Additionally, $\beta_{ji}(t)$ is constrained as follows

$$\beta_{ji,\text{min}} \leq \beta_{ji}(t) \leq \beta_{ji,\text{max}}$$

(3)

where $\beta_{ji,\text{min}}$, $\beta_{ji,\text{max}}$ are the minimum and maximum permissible entrance rate of vehicles, respectively, and $\beta_{ji,\text{min}} > 0$ to avoid long queues and delays at the perimeter of the network and the boundary of neighborhood reservoirs. Moreover, the following constraints are introduced to preserve overflow phenomena within the reservoirs

$$\sum_{i=1}^{N} \left( \beta_{ji}(t) + \varepsilon_i \right) \leq 1, \quad \forall j = 1, \ldots, N$$

(4)

where $\varepsilon_i > 0$ is a portion of uncontrolled flow that enters reservoir $i$. Finally the accumulation $n_i(t)$ cannot be higher than the maximum accumulation $n_{i,\text{max}}$ for each reservoir $i$

$$0 \leq n_i(t) \leq n_{i,\text{max}}, \quad \forall i = 1, \ldots, N.$$  

(5)

Introducing Equation 2 in Equation 1 we obtain the following nonlinear state equation

$$\frac{dn_i(t)}{dt} = \sum_{j \in S_i} \beta_{ji}(t)O_j(n_j(t)) - O_i(n_i(t)) + d_i(t).$$  

(6)

This nonlinear model may be linearized around some set point $\hat{\beta}_{ji}$, $\hat{n}_i$, and $\hat{d}_i$ that satisfy the steady state version of Equation 6, given by

$$0 = \sum_{j \in S_i} \hat{\beta}_{ji}(t)O_j(\hat{n}_j(t)) - O_i(\hat{n}_i(t)) + \hat{d}_i(t).$$

(7)
Denoting $\Delta x = x - \hat{x}$ analogously for all variables and assuming first-order Taylor approximation, the linearization yields

$$
\Delta \dot{n}_i(t) = \sum_{j \in S_i} \Delta \beta_{ji}(t)O_j(\hat{n}_j(t)) + \sum_{j \in S_i} \hat{\beta}_{ji}(t)\Delta n_j(t) - \Delta n_i(t)O'_j(\hat{n}_j(t)) + \Delta d_i(t).
$$

The system in Equation 8 approximates the original system in Equation 6 when we are near the equilibrium point about which the system was linearized.

Applying Equation 8 to a city partitioned in $N$ reservoirs the following state equation (in vector form) describes the evolution of the system in time

$$
\Delta \dot{n}(t) = F\Delta n(t) + G\Delta \beta(t) + H\Delta d(t)
$$

where $\Delta n \in \mathbb{R}^N$ is the state deviations vector of $\Delta n_i = n_i - \hat{n}_i$ for each reservoir $i = 1, \ldots, N$; $\Delta \beta \in \mathbb{R}^M$ is the control deviations vector of $\Delta \beta_{ji} = \beta_{ji} - \hat{\beta}_{ji}, \forall i = 1, \ldots, N, j \in S_i$; $\Delta d \in \mathbb{R}^N$ is the demand deviations vector of $\Delta d_i = d_i - \hat{d}_i$ for each reservoir $i = 1, \ldots, N$; and $F$, $G$, and $H$ are the state, control, and demand matrices, respectively. In particular, $F \in \mathbb{R}^{N \times N}$ is a square matrix with diagonal elements $F_{ii} = -(1 - \hat{\beta}_{ii}(t))O'_j(\hat{n}_j(t))$ if $i \in S_i$, and $F_{ii} = -O'_j(\hat{n}_j(t))$ otherwise, and off-diagonal elements $F_{ji} = \hat{\beta}_{ji}(t)O'_j(\hat{n}_j(t))$ if $j \in S_i$, and $F_{ji} = 0$ otherwise; $G \in \mathbb{R}^{N \times M}$ is a rectangular matrix, where $M \leq N^2$ (depends on the city partition and the set $S_i, i = 1, \ldots, N$) with elements $G_{ji} = O_j(\hat{n}_j(t))$ if the origin reservoir $j$ is reachable from the destination reservoir $i$, and $G_{ji} = 0$ otherwise; $H$ is an identity square matrix of dimension $N$.

The continuous-time linear state system (Equation 9) of the multi-reservoir city may be directly translated in discrete-time, using Euler first-order time discretization with sample time $T$, as follows

$$
\Delta n(k+1) = A\Delta n(k) + B\Delta \beta(k) + \Delta d(k)
$$

where $k$ is the discrete time index, and $A = e^{FT} \approx (I + \frac{1}{2}AT)(I - \frac{1}{2}AT)^{-1}$, $B = F^{-1}(A - I)G$ (if $F$ is nonsingular) are the state and control matrices of the corresponding discrete-time system. This discrete-time linear model (Equation 10) will be used as a basis for feedback control design in the subsequent sections.

**FEEDBACK REGULATORS FOR PERIMETER CONTROL**

The linear control theory offers a number of methods and theoretical results for feedback regulator design in a systematic and efficient way. Linear control theory has extensively been applied in the transportation area over the past 20 years, e.g. ramp metering (17), route guidance (18), traffic signal control (19, 20, 21). In the sequel, we present two alternative optimal control methods that have been used in the past within the Traffic-responsive Urban Control (TUC) strategy for traffic signal control (19), for the design of feedback perimeter control strategies for multi-region and heterogeneously loaded networks. The first methodology is a multivariable feedback regulator derived through the formulation of the problem as a Linear-Quadratic (LQ) optimal control problem. The second methodology that eliminates the need of set values $\hat{\beta}_{ji}$ is obtained through the formulation of the problem as a Linear-Quadratic-Integral (LQI) optimal control problem.
**Perimeter Control Objectives**

In the case of a single-reservoir system (Figure 1(a)) which approximately exhibit an MFD, a suitable control objective is to minimize the total time that vehicles spend in the system including both time waiting to enter and time traveling in the network. It is known that the corresponding optimal policy is to allow as many vehicles to enter the network as possible without ever allowing the accumulation to reach states in the congested regime (13). However, in the case of a multi-reservoir system (Figure 1(b)), such a policy may induce uneven distribution of vehicles in the reservoirs, and, as a consequence, may invalidate the homogeneity assumption of traffic loads within the reservoirs and degrade the total network throughput and efficiency. From this, it can be conjectured (and also shown later in the paper) that the critical accumulation $\bar{n}$ and the maximum throughput $O(\bar{n})$ of a city modeled as a single-reservoir system are very different from the critical accumulation $\bar{n}_i$, $i = 1, \ldots, N$ and the maximum throughput $O_i(\bar{n}_i)$, $i = 1, \ldots, N$ of the same city partitioned in $N$ reservoirs. Moreover, the time each of the reservoirs reach the congested regime is very different.

With these observations at hand, a suitable control objective for a multi-reservoir system aims at: (a) distributing the accumulation of vehicles $n_i$ in each reservoir $i$ as homogeneously as possible over time and the network reservoirs, and (b) maintaining the rate of vehicles $\beta_{ji}$ that are allowed to enter each reservoir around a set (desired) point $\hat{\beta}_{ji}$ while the system’s throughput is maximized. A possible way to act in the sense of point (a) is to equalize the distribution of the relative accumulation of vehicles $n_i/n_{i,\text{max}}$ despite inhomogeneous time and space distribution of arrival flows. Requirement (b) is taken by setting the desired point $\hat{\beta}_{ji}$ be equal to the rate of vehicles correspond to output $O_j(\bar{n}_j)$, $i = 1, \ldots, N$, $j \in S_i$.

The specification of set points $\bar{n}_i$ and corresponding $\hat{\beta}_{ji}$ for monocentric networks with well-defined destination attractions is easy, while heterogeneous networks with multiple regions of attraction would require a non-trivial choice of $\bar{n}_i$. Physically speaking, if a control approach can keep all regions below or close to the critical accumulation of each MFD, $\bar{n}_i$ that maximizes the network outflow, then the problem is well resolved. A big challenge, which will be investigated in the future, is the dynamic choice of $\bar{n}_i$ as a functions of the level of congestion in each region, $n_i(k)$ and the distribution of destinations across the network. For example if heavily directional flows from the periphery of a city pass through a small region to enter the center, the set point $\bar{n}_i$ for the small region should be smaller than the set point $\bar{n}_j$ of the periphery.

**Multivariable Feedback Regulator**

A first approach towards feedback perimeter control based on the dynamics for a city partitioned in $N$ reservoirs (Equation 10) and the control objective mentioned in the previous section is derived as follows. We consider the following quadratic cost criterion:

$$ J = \frac{1}{2} \sum_{k=0}^{\infty} \left( \| \Delta n(k) \|^2_Q + \| \Delta \beta(k) \|^2_R \right) $$

where $Q$ and $R$ are non-negative definite, diagonal weighting matrices. The first term in Equation 11 is responsible for minimization and balancing of the relative accumulation of vehicles $n_i/n_{i,\text{max}}$ in each reservoir $i$. To this end, the diagonal elements of $Q$ are set equal to the inverses of the maximum accumulation of the corresponding reservoirs (see (19, 21) for details). The second
term in Equation 11 is responsible for objective (b) in the previous section and the choice of the weighting matrix \( R = rI \) can be influence the magnitude of the control actions.

Minimization of the performance criterion in Equation 11 subject to Equation 10 (assuming \( \Delta d(k) = 0 \)) leads to the LQ multivariable feedback regulator (22)

\[
\beta(k) = \hat{\beta} - K [n(k) - \hat{n}]
\]

where \( K \) is the steady-state solution of the corresponding Riccati equation which depends only upon the problem matrices \( A, B, Q, \) and \( R \). Note that the corresponding discrete-time linear system (Equation 10) is controllable and reachable and as a consequence a dead-beat gain \( K \) can be off-line calculated for a low value of the scalar weight \( r \).

**Multivariable Integral Feedback Regulator**

The basic approach in integral feedback control is to create a state within the controller that computes the integral of the error signal, which is then used as a feedback term to provide zero steady-state error. We do this by augmenting the description of the original system (Equation 10) with a new state given by

\[
y(k+1) = y(k) + Y \Delta n(k)
\]

where \( y \in \mathbb{R}^m \) is the integral vector and \( Y \in \mathbb{R}^{m \times N} \). The matrix \( Y \) typically consists of 0’s and 1’s such that \( m \) components (or linear combinations of components) of accumulation of vehicles are integrated in Equation 10. The augmented discrete-time model can be written as

\[
\Delta \tilde{n}(k+1) = \tilde{A} \Delta \tilde{n}(k) + \tilde{B} \Delta \beta(k) + \tilde{H} \Delta d(k)
\]

where \( \tilde{n}(k) = [n(k) \ y(k)]^T \) is the augmented state vector, and \( \tilde{A}, \tilde{B}, \tilde{H} \) are the augmented state, control, and demand matrices, respectively. For deriving the integral feedback regulator, the control goal is to minimize the augmented quadratic criterion

\[
J = \frac{1}{2} \sum_{k=0}^{\infty} \left( \| \Delta n(k) \|^2_Q + \| y(k) \|^2_S + \| \Delta \beta(k) \|^2_R \right)
\]

where \( S = sI, s \) positive scalar, is an additional non-negative definite, diagonal weighting matrix.

Considering the augmented cost criterion (Equation 15) and discrete-time model (Equation 14) we obtain the following augmented matrices

\[
\tilde{A} = \begin{bmatrix} A & 0 \\ Y & I \end{bmatrix} \quad \tilde{B} = \begin{bmatrix} B \\ 0 \end{bmatrix} \quad \tilde{H} = \begin{bmatrix} H & 0 \\ 0 & 0 \end{bmatrix} \quad \tilde{Q} = \begin{bmatrix} Q & 0 \\ 0 & S \end{bmatrix} \quad \tilde{R} = R.
\]

Minimization of the performance criterion in Equation 15 subject to Equation 14 (assuming \( \Delta d(k) = 0 \)) leads to the LQI multivariable feedback regulator (22)

\[
\Delta \beta(k) = -\tilde{K} \begin{bmatrix} \Delta n(k) \\ y(k) \end{bmatrix}
\]
where \( \hat{K} \) is the steady-state solution of the corresponding Riccati equation which depends only upon the augmented matrices \( \tilde{A}, \tilde{B}, \tilde{Q}, \) and \( \tilde{R} \). Decomposing \( K = [K_1 \ K_2] \), we get after some algebra the final multivariable integral feedback regulator

\[
\beta(k) = \beta(k-1) - K_p [n(k) - n(k-1)] - K_I [n(k) - \hat{n}]
\]

where \( K_p = K_1 - K_2 \tilde{Y} \) and \( K_I = K_2 \tilde{Y} \) are the proportional and integral gains, respectively. The calculation of \( K \) via solution of the discrete-time Riccati equation, is straightforward and the required computational effort is low even for large-scale cities partitioned in many reservoirs. Moreover, this computational effort is required only off-line, while on-line (i.e. in real-time) the calculations are limited to the execution of Equation 18 with a given constant control matrix \( K \) and state measurements \( n(k) \).

**Constraints and Implementation Issues**

A potential disadvantage of the linear-quadratic theory is that it does not allow for direct consideration of the inequality constraints (Equations 3-5). In this work, the control constraints (Equation 3) are imposed after application of the feedback regulators (Equation 12 or Equation 18) as we will see later. Regarding the state constraints (Equation 5), one may see that the balancing of the relative accumulation of vehicles \((n_i - \hat{n}_i)^2/n_i,max\) via the control objective (Equation 11 or 15) reduces the risk of a reservoir to reach the congested regime in an indirect way. Finally, the overflow constraints (Equation 4) can be satisfied by appropriate selection of \( \beta_{ji,min}, \beta_{ji,max} \). Alternatively, one can solve the same problem as a Model-Predictive perimeter Control problem including all constraints by using the current state (current estimates of the accumulation in each reservoir) of the traffic system as the initial state \( n(0) \) as well as predicted demand flows \( d(k) \) over the a finite-time horizon \((15)\). However, MPC requires that future demands be predicted, and is thus impractical for on-line use.

For the application of the proposed perimeter control strategies to a city partitioned in \( N \) reservoirs, the following parameters should be specified:

- the number of reservoirs \( N \); the set \( S_i \) of origin reservoirs which are reachable from reservoir \( i \)
- the set (desired) accumulation \( \hat{n}_i \) and the corresponding \( O(\hat{n}_i), O'(\hat{n}_i), i = 1, \ldots, N \)
- the set (desired) rate \( \beta_{ji}, \forall i = 1, \ldots, N, j \in S_i \) (corresponding to \( O(\hat{n}_j) \))
- for the LQ feedback regulator (Equation 12) the gain matrix \( K \)
- for the LQI feedback regulator (Equation 18) the design matrix \( \tilde{Y} \), and the gain matrices \( K_p, K_I \)
- the critical accumulation of vehicles for activating \( n_i,act \) and deactivating \( n_i,stop \) the perimeter control strategy, for each reservoir \( i = 1, \ldots, N \); these thresholds should be selected slightly lower than the set accumulation \( \hat{n}_i, i = 1, \ldots, N \)
- the number of signalized intersections located on the boundary between neighborhood reservoirs or the perimeter of the test network for application of the \( \beta_{ji}, \forall i = 1, \ldots, N, j \in S_i \).

The feedback regulators (Equation 12 or Equation 18) are activated in real-time at each sample interval \( T \) and only within specific time windows (e.g. by use of the thresholds \( n_i,act \) and \( n_i,stop \)), based on the current accumulation \( n(k) \) (network load), to calculate the rates of vehicles \( \beta(k) \) to be allowed to enter each reservoir. The required real-time information on the vehicle accumulation \( n(k) \) can be directly obtained via loop detector time-occupancy measurements. The loop detectors may be placed anywhere within the link, but the estimation is most accurate for detector locations around the middle of the link \((19)\).
After the application of the LQ feedback regulator (Equation 12) or the LQI feedback regulator (Equation 18), if the ordered value $\beta(k)$ violates the operational constraints (Equation 3), it should be appropriately adjusted to become feasible. Moreover, the values of $\beta(k - 1)$ used on the right-hand side of Equation 18, should be the bounded values of the previous time step (i.e. after the application of any constraints that may apply) to avoid possible wind-up phenomena in the regulator. The obtained $\beta_{ji}(k)$ values are then converted to arriving flows (by multiplying $\beta_{ji}(k)$ by $O_{ji}(n_{ji}(k))$) and used to define the green periods of the signalized intersections located at the boundary of neighborhood reservoirs or the perimeter of the network. To this end, the latter flows are equally distributed to the corresponding intersections and converted to a (entrance) link green stage duration with respect to the saturation flow of the link and the cycle time of the intersections.

IMPLEMENTATION

Network Description and Simulation Setup

The test site is a 2.5 square mile area of Downtown San Francisco (Financial District and South of Market Area), including about 100 intersections and 400 links with lengths varying from 400 to 1300 feet. The number of lanes for through traffic varies from 2 to 5 lanes and the free flow speed is 30 miles per hour. Traffic signals are all multiphase fixed-time operating on a common cycle length of 90 s for the west boundary of the area (The Embarcadero) and 60 s for the rest.

For the simulation tests, the test area of Downtown San Francisco is modeled via the AIM-SUN microscopic simulator and typical loop-detectors have been installed around the middle of each network link, according to Figure 2(a). The simulation step for the microscopic simulation model of the test site, was set to 0.5 s. For a 4-hours (9:00–12:00) time-dependent scenario with strong demand, ten replications (with different seeds) were carried out with AIMSUN to account for stochastic effects of the simulator. During this simulation scenario the network is filled and severe congestion is faced for 2 hours with many link queues spilling back into upstream links. To derive the MFD of the test site, initially, simulations are preformed with a field-applied, fixed-time signal control plan.

For the application of the proposed perimeter control strategy, the test site is partitioned into three homogeneous reservoirs ($N = 3$) with small variances of link densities (16), according to Figure 2(b). The three reservoirs are separated by blue lines in Figure 2(b), and consist of 112 (yellow colored area), 128 (red colored area), and 147 links (green colored area), respectively. Note that, the total number of links in the network may be different than the sum of links of the three reservoirs for a number of reasons, including not considered links on the boundary of neighborhood reservoirs and origin links on the perimeter of the network. To derive and investigate the shape of the MFDs of the three reservoirs, simulations are performed with the mentioned fixed-time plan. Based on the derived MFDs the LQ feedback regulator (Equation 12) is designed. Then, the perimeter control strategy is applied every $T = 180$ s, a control interval that is twofold or threefold to the cycle length of all the considered intersections. Finally, the perimeter and boundary controllers resulting from Equation 12 are modified to satisfy the control constraints (Equation 3). These settings are then forwarded to 25 signalized intersections located on the boundary of neighborhood reservoirs or the perimeter of the test network for application, i.e. by modifying the green duration of the phases where perimeter and boundary arriving flows are involved.
Macroscopic Fundamental Diagrams

In the sequel, we first derive the MFD for the test network. Then we derive and investigate the shape of the MFDs of the three reservoirs. Based on the obtained MFDs we then design one of the two multivariable feedback regulators (due to space limitations), namely the LQ feedback regulator (Equation 12).

Figure 3(a) displays the MFD resulting for the considered demand scenario and ten replications (R1 to R10), each with different seed in AIMSUN. This figure plots the throughput-load relationship (veh/h vs. veh) in the network for the whole simulation time period. Each measurement point in the diagram corresponds to 180 s. As a first remark, Figure 3(a) confirms the existence of an MFD for the test area of Downtown San Francisco with moderate scatter across different replications. It can be seen that the maximum throughput values (around $30 \cdot 10^4$ veh/h) in Figure 3(a) occur in an accumulation range from 4000 to 6000 vehs. If the accumulation is allowed to increase to values of $n > 6000$ veh, then the network becomes severely congested with states in the congested regime of the MFD, the throughput decreases with accumulation (negative slope) and the system can lead to network-wide gridlock. It is known for a single-reservoir system, as mentioned earlier, that in order to prevent this throughput degradation, the accumulation $n$ should be maintained in the mentioned observed range (close to the critical accumulation $\tilde{n} \approx 6000$ veh) during the heart of the rush while the system’s throughput is maximized (13, 14).
Aboudolas and Geroliminis

Figure 3(b) displays the MFDs of the three reservoirs resulting for the considered demand scenario and ten replications, each with different seed in AIMSUN. It can be seen that all three regions experience MFD with quite moderate scatter across different replications. Nevertheless, there is a clear distinction between congested and uncongested regime for all reservoirs. Note that the time and the accumulation $n_i$ each of the reservoirs reaches the congested regime are very different. The reservoir 3 (red curve) reaches congestion at time 10:30 (for an accumulation $n_3 \approx 1500$ veh) and then it propagates in the reservoirs 2 (green curve) and 1 (blue curve), at time 10:45 (for $n_2 \approx 2000$ veh) and 11:00 (for $n_1 \approx 750$ veh), respectively. This propagation of congestion would not be observable by looking at the unified MFD in Fig. 3(a), which reaches the congestion at time 10:45 for an accumulation $n \approx 6000$ veh. This establishes our conjecture stated in control objective section (see page 6). Note as well that the maximum achievable throughput (flow) is different for each reservoir $i = 1, 2, 3$ (around $7 \cdot 10^4$ veh/h, $11 \cdot 10^4$ veh/h, and $8 \cdot 10^4$ veh/h,
respectively) and occur in accumulation ranges \( n_1 = [500, 1000] \) veh, \( n_2 = [1100, 2250] \) veh, and \( n_3 = [1000, 1700] \) veh, respectively. The difference in maximum flow levels and congested regimes imply corresponding differences of the highest accumulation of vehicles (load) \( n_{i, \text{max}} \) that is reached by each reservoir \( i = 1, 2, 3 \). Thus, reservoir 1 is seen to reach up to 1700 veh, reservoir 2 reaches up to 4250 veh, and reservoir 3 reaches up to 3100 veh. The minimum total flow at maximum load is around \( 2 \cdot 10^4 \) veh/h for all reservoirs.

**Design of the Feedback Perimeter Controller**

The implementation of the proposed perimeter control strategy to the test site corresponds to the suitable design and application of the LQ feedback regulator (Equation 12). A schematic map of the test site partitioned in \( N = 3 \) reservoirs is shown in Figure 2(b). For the proposed partitioning, each reservoir is reachable from the perimeter and the boundary, i.e. \( S = \{1, 2, 3\} \), \( \forall i = 1, 2, 3 \). Thus, each reservoir \( i \) is equipped with one perimeter controller \( \beta_{ii} \) and two boundary controllers \( \beta_{ij}, j \neq i \), and the control vector is given by \( \beta = [\beta_{11} \beta_{21} \beta_{31} \beta_{12} \beta_{22} \beta_{32} \beta_{13} \beta_{23} \beta_{33}]^T \).

The state vector \( n(k) \) includes the accumulation of vehicles for each reservoir \( i \) and is given by \( n = [n_1 \ n_2 \ n_3]^T \).

The set (desired) accumulation \( \hat{n}_i \) for each reservoir \( i \) is selected within the optimal range of the corresponding MFD for maximum throughput, given the analysis in the previous section. More specifically, the following values \( \hat{n}_1 = 600 \) veh, \( \hat{n}_2 = 1250 \) veh, and \( \hat{n}_3 = 1100 \) veh are selected for the current implementation (see Figure 3(b)). The desired rate of the control inputs are based on the corresponding \( O_i(\hat{n}_i) \) values for each reservoir \( i \) and given by \( \dot{\beta} = [0.3 \ 0.2 \ 0.25 \ 0.35 \ 0.25 \ 0.35 \ 0.3 \ 0.2 \ 0.25]^T \). Finally, the minimum and maximum permissible rates are given by \( \beta_{\text{min}} = 0.1^T \) and \( \beta_{\text{max}} = [0.6 \ 0.4 \ 0.5 \ 0.7 \ 0.5 \ 0.7 \ 0.6 \ 0.4 \ 0.5]^T \), respectively.

For the derivation of the gain matrix \( K \in \mathbb{R}^{9 \times 3} \) in Equation 12 it suffices to specify the state matrices \( A \in \mathbb{R}^{3 \times 3}, B \in \mathbb{R}^{3 \times 9} \), and the weighting matrices \( Q \in \mathbb{R}^{3 \times 3}, R \in \mathbb{R}^{9 \times 9} \). The state matrices are developed for the particular network on the basis of the selected set (desired) point \( \hat{n}, \dot{\beta} \) and the linearization according to Equations 8 and 10. The weighting matrices \( Q, R \) in the quadratic cost criterion (Equation 11) are chosen diagonal. More precisely, the diagonal elements of \( Q \) are set equal to the inverses of the maximum accumulation of the corresponding reservoirs, i.e. \( Q_{ii} = 1/n_{i, \text{max}}, i = 1, 2, 3 \) (see page 6). Finally, the diagonal elements of matrix \( R \) were set equal to \( r = 0.00001 \). This low value of the scalar weight \( r \) was found to lead to the following dead-beat gain matrix

\[
K = 10^{-5} \times \begin{bmatrix}
2.57 & 4.09 & 3.30 & 1.28 & 2.04 & 1.65 & 0.93 & 1.48 & 1.19 \\
0.54 & 0.86 & 0.69 & 2.95 & 4.70 & 3.79 & 0.59 & 0.94 & 0.76 \\
0.64 & 1.02 & 0.82 & 0.94 & 1.49 & 1.21 & 3.36 & 5.35 & 4.32
\end{bmatrix}^T.
\]

**Preliminary Results**

Table 1 displays the obtained results in terms of the performance indices Total Time Spent (TTS), Total Distance Traveled (TDT), Speed, and total number of vehicles that exit the network (Total output) during the whole scenario for the perimeter control and no control cases. This table also displays the virtual waiting queues (in veh) that have been stored at the origin links of the network and the number of vehicles within network links inside the three reservoirs at the end of simulation.
Virtual Queue (t_end) and N total (t_end), respectively, because the network is not empty at the end of the simulation. Thus, speed and TTS were also calculated taking into account the virtual queues. Note that by directly extracting performance measures of speed or TTS from the simulator will underestimate the time spent of gated/controlled vehicles.

As can be seen in Table 1, the perimeter control strategy leads to an improvement of the evaluation criteria compared to no control for the whole network, albeit by different percentages. More specifically, when perimeter control is applied, TTS, TDT, and speed are improved in average by 11.7% (5.7% with virtual queues), 1.8%, and 15.4% (8% with virtual queues), respectively, compared to no control. In contrast, the higher virtual waiting queue in perimeter control compared to no control) indicates that the control action does not create queues that spill back to upstream intersections at the boundary of neighborhood reservoirs (βji, i ≠ j, controllers). The simulation ends with a high demand and congestion in the network. In case a decreasing demand was applied after the end of the simulation to reach non-congested states, the benefits are expected to be much higher.

Figures 4(a), 4(b) display the MFDs of the three reservoirs resulting for the considered demand scenario and one replication when no control and perimeter control are applied, respectively.
FIGURE 4 (a) MFDs of the three reservoirs for one replication with no control, (b) MFDs of the three reservoirs for one replication with perimeter control, (c) Total network flow over time for one replication, (d) Vehicles waiting out of the network over time for one replication.

Clearly, when perimeter control is applied, the three reservoirs remain semi-congested and only some states observed in the congested regime; under no control, the network becomes severely congested with states in the congested regime of the corresponding MFDs. Note that, in absence of perimeter control, the throughput at the end of the simulation is around $2 \cdot 10^4$ veh/h for all reservoirs.

To illustrate the qualitative nature of the perimeter strategy actions, the accumulation and flow of the three reservoirs for one replication are depicted in Figures 5(a)-5(b) and Figures 5(c)-5(d), respectively. Traffic conditions are identical for both control cases up to around 9:15 am, when perimeter control is switched on (due to reservoir 3), as accumulation $n_i$ for each reservoir $i$, albeit at different time, the perimeter strategy (a) limit the rate at which arriving vehicles are allowed to enter the network to keep it from becoming congested, and (b) manage the intertransfers between the reservoirs to respect homogeneity in loads over time and the network reservoirs. Thus, throughput is maintained at high levels within the three reservoirs that is close to the target points $O(\hat{n}_i)$ for each reservoir $i$ (corresponding to $\hat{\beta}$), i.e. around $6 \cdot 10^4$ veh/h,
FIGURE 5 (a) Reservoir accumulation of vehicles over time with no control, (b) Reservoir accumulation of vehicles over time with perimeter control, (c) Reservoir flow over time with no control, (d) Reservoir flow over time with perimeter control.

8.5 · 10^4 veh/h, and 7 · 10^4 veh/h (see Figure 5(d)), respectively, in contrast to the no control case (see Figure 5(c)). Remarkably, the accumulation \( n_i \) of each reservoir \( i \) is not exactly maintained to \( \hat{n}_i \) (see Figure 5(b)) thanks to the selection of the weights \( Q_{ii} = 1/n_{i,max} \) in the control objective (Equation 11).

Figure 4(c), 4(d) show the throughput (total flow of all links in the network) and the virtual waiting queues for the same replication, respectively. Figure 4(c) indicates that the perimeter strategy maintains the overall throughput to high values (via appropriate actions within the reservoirs) during the heart of the rush (after 10:30), compared to the no control case, even if it involves longer waiting queues at the origins of the network (c.f. Figure 4(d), Table 1)). This underlines that appropriate designed perimeter control strategies for multi-reservoir systems might prove beneficial in ameliorating some deficiencies associated with single-reservoir systems (e.g. propagation of congestion).
CONCLUSIONS

In this paper, we addressed the problem of perimeter control for congested cities partitioned in reservoirs. To the best of our knowledge, this is the first approach towards the development of generic, cheap, and efficient perimeter control strategies that appropriately account for the spatial and temporal heterogeneity of congestion between the reservoirs. First, by exploiting the properties of the MFD, we described the dynamics of the rush hour in case of multi-region cities that are not uniformly congested. Motivated by the need to distribute the accumulation of vehicles in each reservoir as homogeneously as possible and maintain the rate of vehicles that are allowed to enter each reservoir around a desired point while the system’s throughput is maximized, we then stated our control objective. In order to provide solutions that can be implemented in real time, we introduced two control strategies for determining the perimeter and boundary controllers, namely multivariable feedback regulator and integral feedback regulator. A key advantage of our approach is that it does not require high computational effort and future demand data if the state can be observed.

We have shown via simulation experiments that an area of the Downtown of San Francisco can be partitioned into three homogeneous regions (reservoirs) that approximately exhibit well-defined MFDs. The derived MFDs were used for the design of the multivariable feedback regulator (Equation 12). In addition, the derived perimeter control strategy was tested in simulation, and it was shown to lead to improvements of the considered performance indexes compared to the no control case. Finally, the impact of the perimeter control actions to the three reservoirs and the whole network was demonstrated by use of the corresponding MFDs.

Our feeling is that these findings are of great importance for the traffic engineering community because the concept of an MFD (a) can be applied for heterogeneously loaded large-scale cities with multiple centers of congestion, if these cities can be partitioned in a small number of homogeneous regions, and (b) can be used towards the development of efficient perimeter control strategies.

Future work will deal with the comparison of the proposed feedback perimeter control strategies with other approaches (e.g., perimeter control for single-reservoir systems (13, 14), model-predictive perimeter control (15)) in simulation, and their application in real-life conditions. Another interesting research direction is dynamic partitioning and control of heterogeneously congested networks (16). Also, a straightforward extension of our research is to include queue constraints at the perimeter of the network (14), imposed by less restrictive traffic flow models.

REFERENCES

5. Godfrey, J. W. The mechanism of a road network. Traffic Engineering and Control, Vol. 11,


