A Hierarchical Ramp Metering Control Scheme for Freeway Networks

A. Kotsialos and M. Papageorgiou

Abstract—A nonlinear model-predictive control approach to the problem of coordinated ramp metering is presented. The previously designed optimal control tool AMOC is used within the framework of a hierarchical control structure which consists of three control layers: the estimation/prediction layer, the optimization layer and the direct control layer. More emphasis is given to the last two layers where the control actions on a network-wide and on a local level, respectively, are decided. The hierarchical control strategy combines AMOC’s coordinated ramp metering control with local feedback (ALINEA) control in an efficient way. Simulation investigations for the Amsterdam ring-road are reported whereby the results are compared with those obtained by applying ALINEA as a stand-alone strategy. It is demonstrated that the proposed control scheme is efficient, fair and real-time feasible.

I. INTRODUCTION

Ramp metering aims at improving the traffic conditions by appropriately regulating the inflow from the on-ramps to the freeway mainstream and is deemed as one of the most effective control measures for freeway network traffic.

One of the most effective local ramp metering strategies is the ALINEA feedback strategy and its variations [12], [13], [16], [17]. For coordinated ramp metering a number of design approaches have been proposed such as multivariable control [2] and optimal control [18], [3], [1], [4]. In [9] a nonlinear optimal control problem formulation combined with a powerful numerical optimization algorithm resulted in the AMOC open-loop control tool that is able to consider coordinated ramp metering, route guidance as well as integrated control combining both control measures. In [5], [6], [7] the results from AMOC’s application to the problem of coordinated ramp metering at the Amsterdam ring-road are presented in detail with special focus on the equity issue. A more detailed overview of ramp metering algorithms may be found in [14].

Due to various inherent uncertainties the open-loop optimal solution becomes suboptimal when directly applied to the freeway traffic process. In this paper, the optimal results are cast in a model-predictive frame and are viewed as targets for local feedback regulators which leads to a hierarchical control structure similar to that proposed in [11], albeit with a more sophisticated optimal control approach.

The rest of this paper is structured as follows. In section II the freeway network traffic flow model used for both simulation and control design purposes is briefly described. In section III the formulation of the optimal control problem for ramp metering is presented. In section IV the hierarchical control structure is described. The results of applying ALINEA, as a stand-alone strategy, and the proposed hierarchical strategy are described in section V. Finally, the conclusions and future work are presented in section VI.

II. TRAFFIC FLOW MODELING

A validated second-order traffic flow model is used for the description of traffic flow on freeway networks and provides the modeling part of the optimal control problem formulation. This means that the same model will be used for the traffic flow simulator (METANET) and for the control strategy (AMOC) (see [10] for details).

The network is represented by a directed graph whereby the links of the graph represent freeway stretches. Each freeway stretch has uniform characteristics, i.e., no on/off-ramps and no major changes in geometry. The nodes of the graph are placed at locations where a major change in road geometry occurs, as well as at junctions, on-ramps, and off-ramps.

The time and space arguments are discretized. The discrete time step is denoted by \( T \) (typically \( T = 5 \ldots 15 \) s). A freeway link \( m \) is divided into \( N_m \) segments of equal length \( L_m \) (typically \( L_m \approx 500 \)m), such that the stability condition \( L_m \geq T \cdot v_{f,m} \) holds, where \( v_{f,m} \) is the free-flow speed of link \( m \). This condition ensures that no vehicle traveling with free speed will pass a segment during one simulation time step. Each segment \( i \) of link \( m \) at time \( t = kT, k = 0, \ldots, K \), where \( K \) is the time horizon, is macroscopically characterized via the following variables: the traffic density \( \rho_{m,i}(k) \) (veh/lane-km) is the number of vehicles in segment \( i \) of link \( m \) at time \( t = kT \) divided by \( L_m \), and by the number of lanes \( \Lambda_m \); the mean speed \( v_{m,i}(k) \) (km/h) is the mean speed of the vehicles included in segment \( i \) of link \( m \) at time \( kT \); and the traffic volume or flow \( q_{m,i}(k) \) (veh/h) is the number of vehicles leaving segment \( i \) of link \( m \) during the time period \( [kT, (k+1)T] \), divided by \( T \). The evolution of traffic state in each segment is described by use of the interconnected state equations for the density and mean speed, respectively, [5], [6], [7].

For origin links, i.e., links that receive traffic demand \( d_o \) and forward it into the freeway network, a simple queue model is used (Fig. 1). The outflow \( q_o \) of an origin link \( o \) depends on the traffic conditions of the corresponding mainstream segment (\( \mu, 1 \)), the ramp’s queue length \( w_o \) (veh) and the existence of ramp metering control measures.

A. Kotsialos and M. Papageorgiou are with the Dynamic Systems and Simulation Laboratory, Technical University of Crete, 73100 Chania, Greece apple,markos@dssl.tuc.gr
Fig. 1. The origin-link queue model.

If ramp metering is applied, then the outflow $q_o(k)$ that is allowed to leave origin $o$ during period $k$, is a portion $r_o(k)$ of the maximum outflow that would leave $o$ in absence of ramp metering. Thus, $r_o(k) \in [r_{\text{min},o}, 1]$ is the metering rate for the origin link $o$, i.e., a control variable, where $r_{\text{min},o}$ is a minimum admissible value; typically, $r_{\text{min},o} > 0$ is chosen so as to avoid ramp closure. If $r_o(k) = 1$, no ramp metering is applied, else $r_o(k) < 1$. A similar approach applies to freeway-to-freeway (ftf) interchanges. The evolution of the origin queue $w_o$ is described by an additional state equation (conservation of vehicles). Note that the freeway flow $q_{\mu,1}$ in merge segments is maximized if the corresponding density $\rho_{\mu,1}$ takes values near the critical density $\rho_{\mu,cr}$.

Freeway bifurcations and junctions (including on-ramps and off-ramps) are represented by nodes. Traffic enters a node $n$ through a number of input links and is distributed to the output links. The percentage of the total inflow at a bifurcation node $n$ that leaves via the outlink $m$ is the turning rate $\beta_{nm}$.

III. FORMULATION OF THE OPTIMAL CONTROL PROBLEM

The overall network model has the state-space form

$$ x(k+1) = f[x(k), r(k), d(k)] $$

The state of the traffic flow process is described by the state vector $x \in \mathbb{R}^N$ and its evolution depends on the system dynamics and the input variables. The input variables are distinguished into control variables $u \in \mathbb{R}^M$ and the uncontrollable external disturbances $d \in \mathbb{R}^D$. In this case the state vector $x$ consists of the densities $\rho_{m,i}$, the mean speeds $v_{m,i}$ of every segment $i$ of every link $m$, and the queues $w_o$ of every origin $o$. The control vector $u$ consists of the ramp metering rates $r_o$ of every on-ramp $o$ under control, with $r_{o,\text{min}} \leq r_o(k) \leq 1$. Finally, the disturbance vector consists of the demands $d_o$ at every origin of the network, and the turning rates $\beta_{nm}$ at the network’s bifurcations. The disturbance trajectories $d(k)$ are assumed known over the time horizon $K_F$. For practical applications, these values may be predicted based on historical data and, if necessary, on real-time estimations, see [19].

The coordinated ramp metering control problem is formulated as a discrete-time dynamic optimal control problem with constrained control variables which can be solved numerically over a given optimization horizon $K_F$ [15]. The chosen cost criterion aims at minimizing the Total Time Spent (TTS) of all vehicles in the network (including the waiting time experienced in the ramp queues). The minimization of TTS is a natural objective for the traffic systems considered here, as it represents the total time spent by all users in the network. Penalty terms are added appropriately to the cost criterion, in order for the solution to comply with the maximum queue constraints. The solution determined from AMOC consists of the optimal ramp metering rate trajectories and the corresponding optimal state trajectory.

IV. HIERARCHICAL CONTROL

The solution provided by AMOC is of an open-loop nature. As a consequence, its direct application may lead to traffic states different than the calculated optimal ones due to errors associated with the system’s initial state estimation $x(0)$, with the prediction of the future disturbances, with the model parameters based on which AMOC determines the optimal solution, as well as errors due to unpredictable incidents in the network.

Since estimation, modeling and prediction errors are inevitable, a receding horizon approach (model-predictive control) is employed to address any mismatch between the predicted and actual system behavior. This approach is suitably extended to the hierarchical control system depicted in Fig. 2, which consists of three layers.

The Estimation/Prediction Layer receives as input historical data, information about incidents and real-time measurements from sensors installed in the freeway network. This information is processed in order to provide the current state
estimation and future predictions of the disturbances to the next layer.

The Optimization Layer (AMOC) considers the current time as \( t = 0 \) and uses the current state estimate as initial condition \( x_0 \). Given the predictions \( d(k), k = 0, \ldots, K_P - 1 \), the optimal control problem is solved delivering the optimal control trajectory (translated into optimal on-ramp outflows) and the corresponding optimal state trajectory. These trajectories are forwarded as input to the decentralized Direct Control Layer, that has the task of realizing the suggested policy.

For each on-ramp \( o \) with merging segment \((\mu, 1)\) (Fig. 1) a local regulator is applied with control sample time \( T_c = z_c T \), \( z_c \in N \), in order to calculate the on-ramp outflow \( q_o^*(k_c) \), where \( k_c = z_c \cdot k \). We define the average quantities \( \bar{\rho}_{\mu, 1}^*(k_c) = \sum_{z_c} \rho_{\mu, 1}^*(z_c)/z_c \) and \( \bar{q}_{\mu, 1}^*(k_c) = \sum_{z_c} q_{\mu, 1}^*(z_c)/z_c \), where the *-index denotes optimal values resulting from AMOC.

We distinguish two cases for later comparison. In the first case, the optimal control trajectories are directly applied to the traffic process, i.e.

\[
q_o^*(k_c) = \bar{q}_{\mu, 1}^*(k_c).
\]

In the second case, the Direct Control Layer is actually introduced. More specifically, the regulators ALINEA and flow-based ALINEA ([12], [16]) are employed as local regulators, while the optimal state trajectory is used to determine the set-points for each particular on-ramp. The ALINEA local regulator with set-point \( \hat{\rho}_{\mu, 1} \) reads

\[
q_o^*(k_c) = q_o^*(k_c - 1) + K_r [\hat{\rho}_{\mu, 1} - \rho_{\mu, 1}(k_c)]
\]  

(3)

with \( K_r \) the feedback gain factor. The flow-based ALINEA with set-point \( \hat{q}_{\mu, 1} \) reads

\[
q_o^*(k_c) = q_o^*(k_c - 1) + K_f [\hat{q}_{\mu, 1} - q_{\mu, 1}(k_c)]
\]  

(4)

with \( K_f \) the feedback gain factor. \( q_o^*(k_c - 1) \) is bounded by the maximum ramp flow \( Q_o \) and a minimum admissible ramp flow \( q_{o, \min}^* \). In order to avoid wind-up, the term \( q_o^*(k_c - 1) \) used in both (3) and (4) is bounded accordingly.

In order to avoid the creation of large ramp queues, a queue control policy is employed in conjunction with every local metering strategy (2), (3), (4). The queue control law takes the form

\[
q_o^w(k_c) = -\frac{1}{T_c} [w_{o, \max} - w(k_c)] + d_o(k_c - 1)
\]  

(5)

where \( w_{o, \max} \) is the maximum admissible ramp queue. The final on-ramp outflow then is

\[
q_o(k_c) = \max \{q_o^*(k_c), q_o^w(k_c)\}
\]  

(6)

For more details on the ALINEA strategy and its variations see [16].

The flows \( \bar{q}_{\mu, 1}^* \) are preferable as set-points for local regulation because they are directly measurable without the uncertainty caused by modelling. However, flows do not uniquely characterize the traffic state, as the same flow may be encountered under non-congested or congested traffic conditions. Hence a flow set-point \( \bar{q}_{\mu, 1}^*(k_c) \) is used (in conjunction with flow-based ALINEA), only if \( \bar{\rho}_{\mu, 1}^*(k_c) \leq \rho_{\mu, cr} \) and \( \bar{q}_{\mu, 1}^*(k_c) \leq 0.9 q_{\mu, cap} \), i.e. only if the optimal flows are well below the critical and congested traffic conditions. If \( \bar{\rho}_{\mu, 1}^*(k_c) \geq \rho_{cr, \mu} \) then ALINEA is applied during the period \( k_c \) with set-point \( \bar{\rho}_{\mu, 1} = \bar{\rho}_{\mu, 1}^*(k_c) \).

In any other case, ALINEA is applied with \( \bar{\rho}_{\mu, 1} = \rho_{cr, \mu} \) so as to guarantee maximum flow even in presence of various mismatches.

The update period or application horizon of the model-predictive control is \( K_A \leq K_P \), after which the optimal control problem is solved again with updated state estimation and the disturbance predictions, thereby closing the control loop of AMOC as in model-predictive control. The control actions will be generally more efficient with increasing \( K_P \) and decreasing \( K_A \).

V. APPLICATION RESULTS

A. The Amsterdam network

For the purposes of our study, the counter-clockwise direction of the A10 freeway, which is about 32 km long, is considered. There are 21 on-ramps on this freeway, including the junctions with the A8, A4, A2, and A1 freeways, and 20 off-ramps, including the connections with A4, A2, A1, and A8. The topological network model may be seen in Fig. 3. It is assumed that ramp metering may be performed at all on-ramps. The model parameters for this network were determined from validation of the network traffic flow model against real data [8].

The ring-road was divided in 76 segments with average length 421m. This means that the state vector is 173-dimensional (including the 21 on-ramp queues). With ramp metering applied to all on-ramps, the control vector is 21-dimensional, while the disturbance vector is 41-dimensional.

B. The no-control case

Using the real time-dependent demand and turning rate trajectories as input to METANET without any control measures, heavy congestion appears in the freeway and
large queues are built in the on-ramps. The density evolution profile is displayed in Fig. 4 and the corresponding queue evolution profile in Fig. 5. The excessive demand coupled with the uncontrolled entrance of drivers into the mainstream causes congestion (Fig. 4). This congestion originates at the junction of A1 with A10 and propagates upstream blocking the A4 and a large part of the A10–West. As a result many vehicles are accumulated in the ftf on-ramp of A4 (i.e. a spillback of the congestion onto the A4 freeway) and in the surrounding on-ramps (Fig. 5). The TTS for this scenario is equal to 14,167 veh·h.

C. Application of ALINEA

In this section the application of the ALINEA strategy to all on-ramps is examined. ALINEA is used as a stand-alone strategy for each on-ramp without any kind of coordination. The set-point for each on-ramp $o$ is set equal to the critical density of the corresponding link $\mu$, i.e. $\hat{\rho}_\mu = \rho_{cr,\mu}$ so as to maximize the local freeway throughput. Two cases are considered with respect to the presence or not of the maximum queue constraint in the sense of (6). In case the maximum queue constraints are active, we will assume that the maximum queue length for the urban on-ramps is 100 veh and for the ftf ramps 200 veh. Furthermore, we assume that there is no re-routing of the drivers towards the surrounding urban network when they are confronted with large queues at the on-ramps.

The application of ALINEA without queue constraints leads to a significant amelioration of the traffic conditions and the TTS is reduced to 7,924 veh·h, which is an improvement of 44% compared to the no-control case. The critical point, however, is in the queue evolution profile, where it may be seen (Fig. 6) that a huge queue is formed at the A1 ftf ramp, that actually prevents A1’s demand from triggering the congestion at the junction of A1 with A10.

When maximum queue constraints are considered in the sense of (6), the application of ALINEA becomes less efficient and the resulting TTS equals 10,478 veh·h, a 26% improvement over the no-control case. The reduction of the strategy’s efficiency is due to the fact that the creation of the large queue in the A1 ftf ramp is not allowed any more, hence a congestion is created there, is propagating upstream and triggers ALINEA action in further upstream ramps (Fig. 7).

D. Application of hierarchical control

First the optimal open-loop solution under the assumption of perfect information with respect to the future disturbances for the entire simulation time is considered. This solution serves as an “upper bound” for the efficiency of the control strategy as it relies on ideal conditions. The TTS in this case becomes 6,974 veh·h, which is a 50.8% improvement over the no control case.

As mentioned in section IV, however, the results obtained by the optimal open-loop control are not realistic because the assumption of perfect knowledge of the future disturbances cannot hold in practice. The hierarchical control proposed is able to cope with this problem by employing the rolling horizon technique. For its application we will use $K_P = 360$ (1 hour) and $K_A = 60$ (10 min). For the purposes of this control scenario, it is assumed that the state
of the system is known exactly when AMOC is applied every 10 minutes, which is a fairly realistic assumption.

With respect to the on-ramp demands, we assume that a fairly good predictor is available. Fig. 8 depicts an example of the actual and predicted demand, for the ftf on-ramp A8. The actual demand is input to METANET while the predicted trajectory is input to AMOC. With respect to the prediction of the turning rates, it is possible, based on historical data, to find a mean value for every turning rate for the considered time period. Thus, while METANET receives as input the real time-dependent turning rates, AMOC uses as input the average turning rates. Finally, we assume that there is no mismatch between the model parameters used by METANET and the corresponding parameters used by AMOC and that there are no incidents in the network.

As mentioned in section IV, there are two cases for the application of AMOC results. In the first case, the optimal ramp flows calculated by AMOC are directly applied to the traffic flow process. In the second case, the ALINEA and flow-based ALINEA strategies are employed. In the first case the TTS becomes equal to 8,267 veh·h, which is a 41.6% improvement over the no-control case and 18.5% worsening compared to the optimal open-loop control. When ALINEA is used at the direct control layer, the TTS becomes equal to 8,086 veh·h, which is a 42.9% improvement over the no-control case and 15.9% larger than the TTS of the optimal open-loop control. The density and queue evolution profiles of the second case, are depicted in Figs. 9 and 10, respectively.

Comparing the on-ramp queue evolution profile with the corresponding profile in the case of ALINEA with queue control, the difference between both control strategies becomes apparent. In the ALINEA case, queues are built in the second half of the simulation horizon, in reaction to the congestion that has been formed. In the hierarchical control case the queues are built early in the simulation time in anticipation of the future congestion. Furthermore, this is done in such a manner that the maximum queue constraints are taken into consideration without serious degradation of the strategy’s efficiency.

E. Equity

The maximum queue constraints may also be used to implicitly address the problem of equity [6]. Fig. 11 depicts the average time spent by a vehicle in the ramp queue plus traveling 6.5 km downstream on the freeway, for the no-control case and the three control scenarios considered. It can be seen that in the no-control case the mean travel time is large at the A10-West as a direct consequence of the created congestion. Without queue control ALINEA reduces the mean time for all on-ramps but for A1, where a large peak appears due to the extended delays in the on-ramp queue (Fig. 6). The introduction of the queue constraints for ALINEA reduces the mean travel time at A1 but leads to significant travel time increases in other upstream on-ramps of A10-South and A10-West. Clearly this is not a fair distribution of the ramp delays required for the amelioration of the traffic conditions. In the case of the hierarchical control strategy, the travel times are significantly lower than for no-control or ALINEA with queue constraints for all on-ramps. The high peaks in A1 and/or in A2 are not present anymore at the expense of a relatively low increase of the travel times of the on-ramps upstream of A1 compared to the case of ALINEA without queue constraints. The hierarchical controller’s distribution of the delays is performed in a more balanced way which is more equitable for the drivers, especially those of A1 and A2.

F. Computation time

In order for the hierarchical control to be applied in the field, the computation time needed for the numerical solution of the associated optimal control problem at each
application must be sufficiently low for the real-time application of this approach to be feasible. The required CPU-time varies from application to application, but generally the algorithm converges very fast to an optimal solution within a few CPU-seconds (1MHz P3 processor with Linux), which proves that the real-time application of the control strategy in the field is feasible even for application periods much shorter than the 10 min employed here.

VI. CONCLUSION AND FUTURE WORK

The results of applying local feedback control and rolling-horizon hierarchical coordinated control to the Amsterdam ring-road have been presented. Uncoordinated local control is quite successful in substantially reducing the TTS and lifting congestion up to a certain degree. However, without consideration of ramp queue constraints, local control leads to strongly inequitable results, while with queue control included the resulting efficiency reduces. The hierarchical control strategy combines the network-wide optimal control with local ramp metering strategies within a rolling horizon framework. As expected, it out-performed the local ramp metering approach in terms of both efficiency and equity.

Future work will be focused on the robustness properties of the proposed hierarchical control structure. More precisely, the impact of the following factors will be studied:

- The length of the optimization and of the application horizons $K_F$ and $K_A$ respectively.
- The demand and turning rates prediction errors.
- The mismatch between the model parameters used by AMOC and the corresponding model parameters used by METANET.
- The occurrence of incidents inside the freeway network.
- The maximum queue lengths.

VII. ACKNOWLEDGMENTS

This work was partly funded by the European Commission in the framework of the project EURAMP (IST-2002-23110). The content of this paper is under the sole responsibility of the authors and in no way represents the views of the European Commission. The authors would like to thank Mr. F. Middelham from AVV-Rijkswaterstaat, The Netherlands, for providing the necessary data for the Amsterdam ring-road.

REFERENCES