Technical report 14-037

Distributed model predictive control of freeway traffic networks: A serial partially cooperative approach

H. Majid, M. Hajiahmadi, B. De Schutter, H. Abouaïssa, and D. Jolly

If you want to cite this report, please use the following reference instead:
Distributed Model Predictive Control of Freeway Traffic Networks: A Serial Partially Cooperative Approach

Hirsh Majid1, Mohammad Hajiahmadi2, Bart De Schutter2, Hassane Abouaïssa3, Daniel Jolly3

Abstract—In this paper, a new distributed model predictive control (MPC) scheme for freeway traffic control is proposed. It is aimed at reducing the communication efforts and the computation times in a large network. This new algorithm can coordinate a large number of on-ramps throughout a freeway network in a partially cooperative scheme. The communication is performed between neighboring on-ramps in a special serial fashion and with three different proposed cooperative schemes. The computation time is much less than that of existing distributed MPC approaches in the literature, while achieving a performance close to the one of the centralized MPC method. To evaluate the performance of the proposed partially cooperative schemes, a freeway network case study is selected and the obtained results show a significant decrease in the total computation time with respect to the fully decentralized one. The obtained results show a significant decrease in the total computation time with respect to the centralized and fully cooperative schemes, while maintaining a close distance to the optimal objective function obtained from the centralized case. Furthermore, the performance of the proposed partially cooperative MPC method is evaluated in the case of incidents in the network.

I. INTRODUCTION

Nowadays, improving the traffic conditions is a challenging problem as the motorways are most of time overused and the extension of the infrastructures and capacities is not always possible. Therefore, transportation researchers have focused on other ways to resolve congestion in traffic networks. One of the solutions is to design and implement intelligent traffic control and management systems.

Feedback control is among the freeway traffic control schemes proposed in the literature [1], [2] and it is mostly used for controlling traffic congestion at local on-ramps by keeping the density of vehicles equal or less than a critical level. But, in case of multiple on-ramps, a simple feedback control law would not be able to deal with problem of reducing the overall travel time. Among the heuristic approaches for control of several on-ramps are the traffic responsive feedback control strategy HERO (HEuristic Ramp metering coOrdination) [3] and the so-called CORDIN algorithm [4]. Recently, the model predictive control (MPC) approach has been used in the transportation framework [5], as a solution for coordination between multiple traffic measures. In [6], an MPC scheme is proposed for control of a small freeway network using coordinated ramp metering and variable speed limit control. However, the main issue with MPC is the computation time required to solve the optimization problems inside the MPC controller. The computational complexity becomes a burden when we have a large network in which there are a large number of control variables and real-time control is essential. Therefore, in this paper we establish a new approach in order to make a trade-off between the computation time and the performance of the control system. A distributed control approach realizes this by coordination between several control agents and by decomposing the main centralized control problem into small problems, each handled by the corresponding agent.

In the freeway traffic control context and for coordination of a large number of on-ramps, there is not much related work in the literature. In [7], a hierarchical control approach is proposed for coordinated ramp metering of freeways networks. The optimization algorithm is solved for the whole network and in a centralized manner. The optimal control inputs are sent as reference trajectories to several local controllers at the lower level. In [8], a comparison between global and local MPC algorithms has been done. The problem of decentralized control and the sub-optimality of the obtained control inputs has been solved by establishing a communication scheme between the neighboring subsystems. In [9], a distributed cooperative algorithm was proposed. In this algorithm, each local MPC controller solves the global optimization problem with the overall objective function (in contrast to the decentralized MPC in which each agent deals with its own local objective function). From a practical point of view, fully distributed control schemes might not be feasible because they mostly require a high level communication between agents.

In this paper, we propose a new distributed control scheme capable of linking a large number of on-ramps throughout a freeway network. It is aimed at avoiding communication delays and large computation times. Each on-ramp is controlled by a local controller. The communication between on-ramp controllers is assumed to be performed in a serial scheme [10], where each on-ramp cooperates with its neighboring on-ramps upstream and downstream. Three configurations would be possible: communicating with the upstream neighbor, with the downstream one, or with both. Overall, we propose a partially cooperative MPC framework for control of a chain of on-ramps communicating in a serial form and using either one of the mentioned configurations.

The rest of the paper is organized as follows. Section II
reviews the macroscopic traffic flow model METANET used to predict the evolution of the states. Section III briefly explains model predictive control in the freeway traffic control framework. In Section IV, multiple communication protocols for distributed implementation of the MPC are presented, and this section also elaborately presents our serial partially cooperative scheme. Finally, simulation results for implementation of several MPC control schemes are provided for a benchmark freeway case study in Section V.

II. MACROSCOPIC TRAFFIC FLOW MODELING

METANET [11] is a 2nd-order deterministic macroscopic modeling tool for simulating traffic flow in freeway networks of arbitrary topology and characteristics, including motorway stretches, bifurcations, on-ramps, and off-ramps. Furthermore, METANET allows taking into account control actions such as ramp metering, variable speed limits, and route guidance [12]. The basic equations used to compute the traffic variables for every segment $i$ of motorway link $m$ are the following (we refer to [11]–[13] for a full description of the METANET), see Fig. 1:

$$\rho_{m,i}(k + 1) = \rho_{m,i}(k) + \frac{T}{L_m \lambda_m} [q_{m,i-1}(k) - q_{m,i}(k)]$$  \hspace{1cm} (1)

$$q_{m,i}(k) = \rho_{m,i}(k)v_{m,i}(k)$$  \hspace{1cm} (2)

$$v_{m,i}(k + 1) = v_{m,i}(k) + \frac{T}{\tau} \{ V(\rho_{m,i}(k)) - v_{m,i}(k) \}$$
$$+ \frac{T}{L_m} v_{m,i}(k) \left[ v_{m,i-1}(k) - v_{m,i}(k) \right]$$
$$- \frac{\nu T}{\tau L_m} \rho_{m,i+1}(k) - \rho_{m,i}(k)$$  \hspace{1cm} (3)

$$V(\rho_{m,i}(k)) = v_{\text{free},m} \cdot \exp \left[ - \frac{1}{a_m} \left( \frac{\rho_{m,i}(k)}{\rho_{c,r,m}} \right)^{a_m} \right]$$  \hspace{1cm} (4)

where $v_{\text{free},m}$ is the free-flow speed of link $m$, $\rho_{c,r,m}$ is the critical density per lane of link $m$, and $a_m$ is a parameter of the fundamental diagram (see also Fig. 2). Furthermore, $\tau$, a time constant, $\nu$, an anticipation coefficient, and $\kappa$ are constant parameters.

In order to take into account the speed decrease caused by merging phenomenon a due to on-ramp, the following term is added to (3):

$$- \frac{\delta T q_o v_{m,1}(k)}{L_m \lambda_m (\rho_{m,1}(k) + \kappa)}$$  \hspace{1cm} (5)

where $\delta$ denotes a model parameter and $v_{m,1}$ and $\rho_{m,1}$ are the speed and the density of the first segment of the link to which the on-ramp is connected. Origins are modeled with a simple queue model: the length of the queue equals the previous queue length plus the demand $d_o(k)$, minus the outflow $q_o(k)$:

$$w_o(k + 1) = w_o(k) + T[d_o(k) - q_o(k)]$$

The outflow $q_o(k)$ of an origin link $o$ depends on the traffic conditions on the main-stream and, for the metered on-ramp, on the ramp metering rate, where $r_o(k) \in [0, 1]$:

$$q_o(k) = \min \left[ d_o(k) + \frac{w_o(k)}{T}, Q_o r_o(k), Q_o \right]$$

where $Q_o$ denotes the on-ramp capacity flow, $\rho_{\text{max},m}$ denotes the maximum density of link $m$, and $\rho_{m,1}$ is the density of the first segment of link $m$ to which on-ramp $o$ is connected.

The upstream speed of the mainstream link is often assumed to be equal to the speed of the first segment, i.e. $v_{m,0} = v_{m,1}$. The downstream density of the mainstream link is often assumed to be equal to the density of the last segment $N$ in free flow, and to be equal to the critical density in congested flow:

$$\rho_{m,N+1}(k) = \begin{cases} \rho_{m,N}(k) & \text{if } \rho_{m,N}(k) < \rho_{c,r,m} \\ \rho_{c,r,m} & \text{if } \rho_{m,N}(k) \geq \rho_{c,r,m} \end{cases}$$  \hspace{1cm} (6)

In the next section, the METANET model is used in the MPC framework for prediction of traffic states.

III. MODEL PREDICTIVE CONTROL

Model Predictive Control (MPC) [14] is an advanced control method recently used for control of traffic networks [6], [15]. The main concept is to use a prediction model of the traffic network and an objective function assessing the desired performance, and to determine the optimal control inputs through solving an optimization problem. Fig. 3 show the MPC scheme for freeway traffic control. In our case, the METANET model is used to predict the evolution of the states of a freeway network over a prediction horizon. The optimization algorithm minimizes the objective function and finds a sequence of optimal control inputs for the whole prediction horizon, but only the first sample of the control input is applied to the traffic network and the procedure is performed in a rolling horizon style (see Fig. 4).
As the objective function, we select the total time spent in the freeway network, composed of the time vehicles spend in queues at on-ramps and the travel time on the freeway itself. The Total Time Spent (TTS) function is formulated as follows:

\[ J_{\text{TTS}}(k_c) = T \sum_{k_c = 1}^{M(k_c + N_p) - 1} \left( \sum_{(m,i) \in I_d} \rho_{m,i}(k) L_m \lambda_m + \sum_{o \in O_{\text{all}}} \omega_o(k) \right) \tag{9} \]

where \( T \) is the simulation sample time, \( k_c \) is the controller time step counter, and \( k \) is the simulation time step counter. In fact, we assume that the controller time step length is an integer multiple of the simulation time step length: \( T_c = MT \). Moreover, \( N_p \) is the control horizon, \( \rho_{m,i} \) is the density of segment \( i \) of link \( m \), \( \omega_o \) is the queue length at origin \( o \), and \( I_d \) and \( O_{\text{all}} \) are the set of all links and segments and the set of all origins, respectively. Furthermore, control inputs that have high fluctuations are not desired. This is due to the fact that in reality traffic signals cannot vary with high frequency over time. Further, high fluctuations in control inputs may cause instability in some cases. Therefore, a penalty term on control input deviations is usually added to the objective function. In our case, the control inputs are the metering signals of on-ramps. The penalty term is formulated as

\[ J_{\text{Pen}}(k_c) = \zeta \sum_{l = k_c}^{k_c + N_p - 1} \sum_{o \in O_{\text{ramp}}} \left| r_o(l) - r_o(l - 1) \right| \tag{10} \]

where \( r_o \) is the metering signal and \( O_{\text{ramp}} \) is the set of indices of metered ramps\(^1\). The \( \zeta \) is a weighting coefficient. Also, to reduce the complexity, control variables are sometimes taken constant after passing a predefined control horizon \( N_c \). Noting this, \( N_p \) in (18) should be replaced by \( N_c \). Finally, the total objective function is the sum of (9) and (10).

IV. COMMUNICATION PROTOCOLS

In this section, we follow an agent-based modeling approach to decompose the centralized model predictive control problem for a large network into smaller problems each solved by local controllers. In the freeway network case, each agent corresponds to a part of the freeway stretch that is composed of multiple segments and links and possibly on-ramps and off-ramps. The agent’s state variables are density, velocity, and flow of vehicles in the multiple segments of the network under control of the agent and the control inputs are the ramp metering rates. Based on the traffic states and the communicated traffic states/control inputs from other agents in the network, an agent computes the optimal (local) ramp metering rates by solving an optimization problem inside the MPC framework. As mentioned in [10], an agent can employ different schemes for communication between other agents and for decision making:

- It can receive information from other agents during the (local) decision making process and send its outputs only afterwards.
- It can receive and send information after all agents have computed their control inputs.
- It can receive information after a neighboring agent has finished its computation process and also send information only after deciding on its own actions.
- It can receive and send information a couple of times before making a decision.

In this paper, the agents communicate with each other serially. This scheme has also been proposed in [17] in which only one agent at a time performs a local optimization step and sends information to a neighboring agent. After that, the neighboring agent performs a local computation step and sends information to the next neighboring agent, and

\(^1\)It should be noted that the outflow from mainstream origins can also be controlled, so in that case they can also be included in the set \( O_{\text{ramp}} \).
so on. In [17], the proposed serial approach is shown to have desired properties in terms of convergence speed and quality of the solution. A disadvantage of this scheme is that agents have to wait for the neighboring agents to solve their optimization problem. In the following, we list one by one all communication schemes that were studied in this work.

A. Centralized MPC

Generally speaking, the best performance can be obtained using the centralized control [8], [9], [18]. However, the centralized control becomes intractable in practice for large networks due to the large computation time required. Here, the control variables and constraints of all the network agents are grouped and the total cost function 9 is minimized for the whole network.

B. Decentralized MPC

In this scenario, the optimization is performed locally for the part network controlled by its associated agent. It means that a local objective function is taken into account in the optimization, which consists of the density and queue length variables of the part of the network assigned to the agent. In this scheme, there is no communication between agents. The controlled system is subject to constraints on the maximum queue length at each on-ramp and the maximum and the minimum ramp metering rate. The total objective function contains two terms for the TTS, one term for the mainstream traffic and one term for the on-ramp queues, and one term added to penalize abrupt variations in ramp metering rates.

C. Full cooperative MPC

In the full cooperative MPC, instead of defining local cost functions, a global cost function for the whole network is defined. The aim of this method is to obtain a good performance as efficiently as the centralized control with a reduced computational time that permits MPC to be applicable in real time. It is assumed that all agents communicate their state and control variables and each minimizes the global cost function. The size of the optimization problem and constraints in this case is significantly less than the centralized case. However, the amount of communication required to coordinate all agents and to reach a consensus in the optimal global cost for the whole network is considerable.

D. Partially cooperative MPC

In this scheme, we aim at reducing the computation time even further while keeping a close relation to the optimal cost achieved by the centralized scheme. One can realize that the results achieved by the fully cooperative MPC is outstanding compared to the decentralized MPC case in terms of performance and compared to the centralized MPC in terms of computational time. However, even with this method, the optimization is time consuming and might not be tractable for large networks mostly because of the high level of communication (required to calculate the global cost function at each control step). Therefore, in our proposed serial partially cooperative scheme, we assume that there is a full communication between only the neighboring agents of the networks. In the freeway network case, the serial communication scheme means that the upstream agent sends its traffic states to the first segment of the downstream agent. At the same time, the downstream agent sends the traffic states of its first segment to the last segment of the upstream agent. Based on this serial communication protocol, three different cooperative schemes are proposed:

- **Upstream scheme**: The cost function of each agent is defined based on its local objective function plus the one of its upstream agent.
- **Downstream scheme**: The cost function of each agent is defined based on its local objective function plus the one of its downstream agent.
- **Upstream/Downstream scheme**: The cost function of each agent is defined based on its local objective function plus the ones of its upstream agent and its downstream agent.

Hence, choosing one of the above schemes, we establish a partially cooperative MPC framework. By partial, we mean that the cost functions of the agents does not include only the local ones but also the ones of the neighboring agents and using the serial communication between all (neighboring) agents of the network, a consensus on the minimized global cost function would be achieved while keeping the communication and computation effort at a low level. In the next section, the performances of the proposed serial partially cooperative schemes along with the conventional centralized and decentralized approaches are compared for a freeway benchmark network.

V. CASE STUDY

The network case study is shown in Fig. 6. The network is a 14 km freeway stretch. The stretch is composed of 7 links and 14 segments, each with 2 lanes denoted by $\lambda = 2$. Moreover, there are seven on-ramps throughout the stretch with equal distance from each other. It is assumed that all segments share identical parameters, as presented in Table I. Seven agents are assigned to the network, each controls one on-ramp. The traffic demands for the mainstream origin and the on-ramps are selected as in Fig. 5, in order to obtain a scenario with a high level of congestion in the network. The simulation period is three hours, which corresponds to 180 controller steps and 1800 simulation steps.

The control and communication schemes presented in Section IV are implemented and the obtained results for the minimum total time spent (TTS) and the total computation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>40 veh/lane/km</td>
<td>$\nu$</td>
<td>60 km/h</td>
</tr>
<tr>
<td>$\rho_{\text{max}}$</td>
<td>180 veh/lane/km</td>
<td>$\delta$</td>
<td>0.0122</td>
</tr>
<tr>
<td>$T$</td>
<td>10 s</td>
<td>$\tau$</td>
<td>18 s</td>
</tr>
<tr>
<td>$\rho_{\text{cr}}$</td>
<td>33.5 veh/lane/km</td>
<td>$\alpha$</td>
<td>0.1</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>1.867</td>
<td>$C_{\text{ramp}}$</td>
<td>0.4</td>
</tr>
<tr>
<td>$C_{\text{min}}$</td>
<td>4200 veh/h</td>
<td>$C_{\text{ramp}}$</td>
<td>2000 veh/h</td>
</tr>
</tbody>
</table>
Fig. 5. Traffic demand in the origin and all on-ramps

TABLE II
RESULTS OF THE IMPLEMENTATION OF THE DIFFERENT SCENARIOS
WITHOUT INCIDENT

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Np</th>
<th>Nc</th>
<th>TTS (veh-h)</th>
<th>CT (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncontrolled system</td>
<td></td>
<td></td>
<td>5394.3</td>
<td>≃ 0</td>
</tr>
<tr>
<td>Decentralized MPC</td>
<td></td>
<td></td>
<td>5347.1</td>
<td>8072</td>
</tr>
<tr>
<td>Partially cooperative MPC:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upstream</td>
<td>7</td>
<td>5</td>
<td>5308.4</td>
<td>8271</td>
</tr>
<tr>
<td>Downstream</td>
<td>7</td>
<td>5</td>
<td>5324.6</td>
<td>8369</td>
</tr>
<tr>
<td>Up/Down-stream</td>
<td>7</td>
<td>5</td>
<td>5318.6</td>
<td>8583</td>
</tr>
<tr>
<td>Fully cooperative MPC</td>
<td>16</td>
<td>12</td>
<td>5222.8</td>
<td>87133</td>
</tr>
<tr>
<td>Centralized MPC</td>
<td>16</td>
<td>12</td>
<td>5202.5</td>
<td>131090</td>
</tr>
</tbody>
</table>

The simulations are performed on a 64-bit Windows PC with a 2.8GHz Intel Core i7 processor and 8Gb RAM

time (CT) are presented in Table II. For solving the optimization problems inside the MPC framework, the \texttt{fmincon} function of MATLAB \(^2\) is used together with a multi-start strategy for random initial points in order to avoid reaching local optimal points only.

Generally, the best performance of the systems is obtained using the centralized control. However, as mentioned in Section IV, the required computation time is very high for large networks (see the result for the centralized MPC in Table II).

As for the decentralized scheme, for each on-ramp and the associated 2 km segment, the optimization is performed locally. Since it is assumed that agents do not communicate and do not consider the overall cost function, the obtained value for the total TTS is higher than in the centralized case.

In the full cooperative MPC, each agent instead of the local cost function, uses the global cost function of the whole network and full communication between all agents is necessary. The TTS performance is close to the centralized case but the computation time is relatively high, which makes real-time control of the freeway impossible.

Finally, as for our proposed partially cooperative MPC scheme, 3 main configurations are proposed as described in Section IV and as illustrated in Fig. 7. For instance in the Upstream case, the objective function for agent 2 is calculated based on the travel times of links 1 and 2, see Fig. 7-top. The results of employing the partially cooperative schemes are presented in Table II.

It can be inferred that in all three cases of the partially cooperative schemes, the computation time is reduced significantly compared to the fully cooperative method and meanwhile, the optimized global cost function is kept close to the one of the centralized MPC. Note that for a complicated traffic scenario, we would expect much better performance for the proposed partially cooperative MPC schemes compared to the decentralized case.

A. All scenarios in case of incident

All scenarios have been tested in the case of incidents in two different locations as well as in different times: for agent 2, segment 4, from simulation step 180 to 240 and for agent 6, segment 12, from simulation step 480 to 600. Table III shows the results of the incident scenarios.

From Tables II and III, it can be inferred that in all cases with/without accident, the centralized MPC and the fully cooperative MPC has better performance compared to other scenarios. But, the computation time is very long. In the proposed serial partially cooperative schemes the incidents are handled well and the computation time is reduced significantly, while keeping the performance close to the centralized case.

TABLE III
RESULTS OF THE IMPLEMENTATION OF THE DIFFERENT SCENARIOS
WITH INCIDENT

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Np</th>
<th>Nc</th>
<th>TTS (veh-h)</th>
<th>CT (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncontrolled system</td>
<td></td>
<td></td>
<td>6703.5</td>
<td>≃ 0</td>
</tr>
<tr>
<td>Decentralize system</td>
<td>7</td>
<td>5</td>
<td>6654.1</td>
<td>7945</td>
</tr>
<tr>
<td>Partially cooperative MPC:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upstream</td>
<td>7</td>
<td>5</td>
<td>6604.1</td>
<td>8226</td>
</tr>
<tr>
<td>Downstream</td>
<td>7</td>
<td>5</td>
<td>6586.2</td>
<td>7921</td>
</tr>
<tr>
<td>Up/Down-stream</td>
<td>7</td>
<td>5</td>
<td>6585.3</td>
<td>7747</td>
</tr>
<tr>
<td>Fully cooperative MPC</td>
<td>16</td>
<td>12</td>
<td>6536.2</td>
<td>85969</td>
</tr>
<tr>
<td>Centralized MPC</td>
<td>16</td>
<td>12</td>
<td>6507.6</td>
<td>130675</td>
</tr>
</tbody>
</table>
to the fully cooperative scheme and the centralized case. Moreover, with the proposed partially cooperative schemes, not all the agents in the network need to be connected to each other and therefore the communication effort is also reduced significantly. This is very important for large networks where it is difficult to make all agents communicate in real time. Finally, Fig. 8 shows the densities of all segments for the whole simulation period, in the absence of incidents. It can be observed, that the congestion level and duration is lower in the fully cooperative MPC method. In the partially cooperative case, the duration of the congestion is reduced compared to the decentralized case, and furthermore, the free flow conditions for all segments is achieved is shorter time.

VI. CONCLUSION

A new distributed MPC scheme the serial partially cooperative MPC has been proposed. The computation time and the amount of communication required for our proposed framework is significantly less than the ones for the centralized and fully distributed schemes. Furthermore, in terms of the total cost, our proposed schemes achieve relatively close values to the optimal cost obtained from the centralized scheme. As future extension of the current work, we will further validate our methodology using different traffic scenarios as well as by using micro-simulation to simulate the traffic network. Moreover, we will integrate this scheme with the so-called event-triggered concept in the sense that solving local optimization problems may be skipped at some specific time instants, in which the network outflow is an acceptable level. Integrating the event-triggered concept into our partially cooperative schemes, the total computation time would be reduced even further.

ACKNOWLEDGMENT

This work is the result of a short term research visit at the Delft Center for Systems and Control (DCSC), Delft University of Technology, The Netherlands, supported by the Cost Action TUD1102.

REFERENCES