Distributed Optimal Control of Smart Electricity Grids with Congestion Management

D. Bao Nguyen, Jacquелиen M. A. Scherpen, Senior Member, IEEE, and Frits Bliek

Abstract—In this paper, we consider the balancing problem in a hierarchical market-based structure for smart energy grids that is based on the Universal Smart Energy Framework. The large-scale introduction of renewable, intermittent energy sources in the power system can create a mismatch between the forecasted (day-ahead) and the actual supply and demand. Without a proper control strategy, this deviation could lead to network overloads and commercial losses. We present a multi-level distributed optimal control formulation to the problem, in which the appliances of prosumers that can provide flexibility are optimally dispatched based on local information. The control strategy takes the capacity limitations of the distribution network into account. We provide example simulation results, obtained by distributed model predictive control.

Note to Practitioners—We propose a control strategy that aims to minimize the imbalance between forecasted and actual supply and demand in electricity grids. This is important, because the imbalance can lead to commercial losses for the stakeholders. Since the number of agents (i.e., households) in the power network is typically large, centralized controllers are not feasible due to scalability issues. We instead develop a distributed controller that solves the problem using only local information. We demonstrate our algorithm through simulations, which are implemented on a single computer. In practice, households can have smart meters on which the individual controllers run, thereby obtaining the solution in a parallel fashion.

Index Terms—Optimal control, multi-level distributed control, smart grid, Universal Smart Energy Framework.

NOMENCLATURE

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>Power consumption/production of device.</td>
</tr>
<tr>
<td>η, C</td>
<td>Conversion factors.</td>
</tr>
<tr>
<td>q</td>
<td>Heat demand of prosumer.</td>
</tr>
<tr>
<td>f_on, f_off</td>
<td>Number of time-steps a device has been on/off.</td>
</tr>
<tr>
<td>T_on, T_off</td>
<td>Bounds on f_on, f_off (min and max).</td>
</tr>
<tr>
<td>A</td>
<td>Information sharing matrix.</td>
</tr>
<tr>
<td>goal_i</td>
<td>Goal function, the DAP share of prosumer i.</td>
</tr>
<tr>
<td>x_i</td>
<td>Real (physical imbalance) of prosumer i.</td>
</tr>
<tr>
<td>z</td>
<td>Aggregator index.</td>
</tr>
<tr>
<td>N</td>
<td>Total number of prosumers.</td>
</tr>
<tr>
<td>n</td>
<td>Number of prosumers per aggregator.</td>
</tr>
<tr>
<td>f_i, g_i</td>
<td>Flexible and fixed load of prosumer i.</td>
</tr>
<tr>
<td>L_max</td>
<td>Distribution network capacity limit.</td>
</tr>
<tr>
<td>λ_i, µ</td>
<td>Lagrangian multipliers.</td>
</tr>
<tr>
<td>β, γ</td>
<td>Subgradient iteration step sizes.</td>
</tr>
<tr>
<td>ε</td>
<td>Subgradient iteration stopping criterion.</td>
</tr>
</tbody>
</table>

I. INTRODUCTION

Environmental concerns and changes in power usage have led to the emergence of smart grids. There is a drive to reduce CO₂ emission and to turn towards renewable energy sources (e.g., solar energy, wind energy, biomass). The European Union has set targets of (1) reducing greenhouse gas emission by 20% relative to the 1990 level and (2) each member state achieving a 20% share of energy consumption from renewable sources; a policy to be realized by 2020 [1]. However, these energy sources are characterized by intermittency: the production depends heavily on weather conditions. End-users, who were traditionally consumers, can become producers too by using, for example, photovoltaic solar panels or µCHP (micro combined heat and power) devices. They are henceforth called prosumers.

The need to accommodate fluctuating generation while avoiding network overloads creates an optimization problem: what is the optimal way to supply the required power demand, while compensating at the same time for (short-term) deviations between the forecasted and the actual supply and demand of power in the system? Since currently there is no economically efficient way to store electricity in large quantities, these deviations have to be canceled out to maintain the overall system balance and make optimal use of the renewable power generation. To overcome this problem, smart grids exploit the flexibility of appliances; the combined flexibility of the network of households can be used to optimize the performance of the energy system. The contribution to the balancing problem from the consumer side is often referred to as demand
response. Other possibilities include utilizing interacting grids for storage, for example, Power-to-Gas facilities [2].

A natural way to approach the problem is to use model predictive control (MPC) [3], as it enables the incorporation of future, weather-dependent predictions in the decision process. Examples of MPC application to smart grids include [4], [5], and [6]. Giselsson and Rantzer [7] suggest a distributed version of this technique, in which agents make their own decisions relying only on local information. The distributed formulation is obtained via dual decomposition and Lagrangian relaxation [8], [9]. Larsen et al. [10] apply the strategy to control a network of households with washing machines (flexible consumption). Distributed MPC is then implemented to balance between heat demand and supply in a network of households with $\mu$CHPs (flexible production) [11]. In both cases, the households are connected using an information sharing model that is introduced in [12]. Biegel et al. [13] propose a control method based on dual decomposition to achieve congestion management. However, their study is limited to one level of hierarchy, and only deals with flexible consumption. Various multi-level distributed MPC schemes, but without congestion management, are described in [14], [15], and [16]. We aim to combine all aforementioned efforts into one model.

The main contribution of this work is to build up on [11] (where only the electricity production is flexible) and [12], and consider the scenario where the prosumers (households) have $\mu$CHPs and heat pumps, i.e., both flexible production and flexible consumption is present in the same setting. The $\mu$CHP and heat pump are both connected to heat buffers that can store heat converted from surplus electricity. Furthermore, we embed our distributed MPC controller in the Universal Smart Energy Framework, in which there is also an aggregator level above the prosumer level, and the two levels are coupled through a goal function. The objective is to minimize the prediction error between the forecasted (represented by the goal function) and the actual supply and demand in the system, by utilizing the flexible appliances of the households. The deviation we treat can arise from the forecasting inaccuracies of both flexible loads (e.g., $\mu$CHPs, heat pumps) and fixed loads (e.g., solar panels, TVs). While doing this, we also take measures to avoid overloading the distribution network. Our control method handles two different Lagrange dual variables, associated with two different type of constraints. The coupling constraint between the prosumers can be relaxed such that a distributed formulation among them is obtained, whereas the $\mu$DSO constraint requires a central coordinator, resulting in a multi-level distributed optimization problem. A preliminary version of this research is reported in [17]. Compared to that report, here we also describe a method to quantify flexibility, develop our model to a multiple-aggregator-per-transformer case, and provide extended simulations. Additionally, we elaborate on the simulation results in more detail.

The rest of the document is organized as follows. First, we introduce the Universal Smart Energy Framework, and describe our problem within its hierarchical market-based structure in Section II. Sections III-V develop the distributed optimal control scheme for the balancing problem. We then present our implementation for three different scenarios, and the corresponding result analysis in Section VI. We end with our conclusions in Section VII.

II. PROBLEM SETTING

A. Framework

The Universal Smart Energy Framework (USEF) [18] is an initiative by a collective of top sector companies to standardize smart grid solutions for the European energy market. Their aim is to create a platform to drive the fastest, most cost-effective route to an integrated smart energy future. USEF delivers a common standard on which to build all smart energy products and services. It unlocks the value of flexibility by making it a tradable commodity, and delivering a market structure, associated rules, and tools to make it work effectively. Flexibility can be invoked for grid capacity management to avoid or reduce peak loads, and allows for active balancing through optimization between supply and demand. The framework is designed to offer fair market access and benefits to the stakeholders, and is accessible to anyone internationally.

In this study, we treat the following USEF stakeholders: the Balance Responsible Party (BRP), aggregators, prosumers, and the Distribution System Operator (DSO). Electricity is traded between the suppliers and the BRPs over the wholesale energy market (day-ahead) and imbalance market (operation time). The BRPs dispatch the electricity to the aggregators, which in turn deliver to the prosumers. The aggregator is a new stakeholder in energy grids that groups the prosumers into clusters. Its responsibility is to accumulate and offer flexibility on behalf of the connected prosumers, with the aim of maximizing the value of flexibility. The DSO is responsible for the distribution of power and to resolve any disturbances that might interfere with that task. In this context, the main task of the DSO is to detect and resolve any congestion that might occur in the distribution lines.

USEF employs a market-based control mechanism which consists of five phases: contract, plan, validate, operate, and settlement. Contractual agreements between the stakeholders are established in the first phase. In the plan phase, a day-ahead forecast of the energy consumption is made, which is then validated by the DSO in the validate phase. The two phases are iterated until an agreement is reached on the forecast. In the operate phase the system aims to follow the plan that has been created in the first two phases, and balances between the forecast and actual electricity load by procuring flexibility. Financial reconciliation is completed in the settlement phase. An overview of the USEF structure and market-based control mechanism is shown in Fig. 1. Parts that are not relevant to this research are omitted from the figure, for full details, see [18].

B. Problem statement

The work presented here is focusing on the operate phase of USEF, with a layout as seen in Fig. 2. Note, that compared to [17], we now look into the case where multiple aggregators are constrained by the same distribution network capacity limit (i.e., all prosumers are connected to the same transformer which couples them).
We assume that the electricity portfolio has already been forecasted and agreed on by the aggregators, BRP, and DSO (i.e., it is free of congestion). This electricity portfolio is referred to as the day-ahead planning. Given prosumer heat demands, the goal is to fulfill that demand while keeping the electricity load as close to the day-ahead planning as possible (i.e., minimize the prediction error). For simplicity, we assume that each prosumer is equipped with one appliance, either a \( \mu \)-CHP (representing electricity production) or a heat pump (representing electricity consumption), see Fig. 3. The method described in this paper is also applicable to the case where one prosumer can have multiple flexible devices, for example, by considering each device as an agent.

We propose a hierarchical, three-level structure (BRP, aggregators, prosumers), in which the day-ahead planning is spread over the aggregators, and subsequently the prosumers connected to them, based on the available flexibility of the prosumers. At the level of the prosumers, the minimization of the prediction error in the operate phase is solved in a distributed manner, where each prosumer contributes to the optimization process based only on local information exchange with their neighbors. In [19], which considers a similar distributed MPC framework, although not in a market-based structure, it has been shown that the more information the prosumers share with each other, the less imbalance there is in the network. The role of the DSO is included in the model to avoid violation of the distribution network capacities. To steer the flexible appliances of the prosumers, a goal function is introduced which relates to the day-ahead planning.

Fig. 2. Our model considers the three-level structure within the informational network, as well as the connections to the distribution lines in the physical network. LV, MV, and HV stand for low voltage, medium voltage, and high voltage networks, respectively.

III. PRELIMINARIES

Throughout the paper, we use the convention of assigning a negative sign to electricity production (supply), and a positive sign to consumption (demand). The total load of a household is the signed sum of the two quantities. We use the terms “appliance” and “device” interchangeably.

A. Quantification of flexibility

In our model, flexibility is regarded as the ability to shift the production or consumption of appliances in time, without changing the total energy production or consumption. By utilizing the flexibility of the devices (i.e., turning on/off devices based on the load measured in the network), demand side management can be performed. We describe a method
to quantify the flexibility a thermal appliance can offer at a given time, based on [20]. The flexibility of the \( \mu \text{CHP} \) and heat pump, both in combination with a heat buffer, can be measured as the potential power increase or decrease with respect to a baseline (i.e., current) power production or consumption. By doing so, two scenarios can be distinguished:

- ramp-up flexibility: increasing the electricity consumption of the household by turning off the \( \mu \text{CHP} \) or turning on the heat pump,
- ramp-down flexibility: decreasing the electricity consumption of the household by turning on the \( \mu \text{CHP} \) or turning off the heat pump.

We assume that the water level of the buffer remains constant, and only the heat content is changing. The buffer fills according to

\[
L[\tau] = L_0 + \frac{\eta}{C} \left[ P[\tau] - \sum_{\tau} q[\tau] \right], \quad 0 \leq L[\tau] \leq 1, \tag{1}
\]

where \( L[\tau] \) is the generalized buffer level after \( \tau \) time-steps, \( L_0 \) is the initial buffer level, \( \eta \) is a ratio between electric and thermal power, \( C \) is a conversion factor from thermal power to the buffer level, and \( P \) is the power produced (in case of a \( \mu \text{CHP} \)) or consumed (in case of a heat pump) while filling the buffer. The household has a heat demand \( q[\tau] \), with which the buffer is drained. Parameters \( \eta, C, \) and \( P \) may vary for the \( \mu \text{CHPs} \) and heat pumps, we denote their specific parameters with subscripts \( C \) and \( H \), respectively.

From Eq. (1) we can derive the remaining available electrical capacity of the heat buffer:

\[
P = \frac{C(1 - L_0) + \sum_{\tau} q[\tau]}{\eta}. \tag{2}
\]

Taking the power limits into account, the ramp-up and ramp-down flexibilities of the \( \mu \text{CHP} \) can be then given by the following expressions:

\[
F^+_{C}[\tau] = \min \left\{ \frac{\delta C C_L - \sum_{\tau} q[\tau]}{\eta C}, -P_C \right\}, \tag{3}
\]

\[
F^+_{C}[\tau] = \max \left\{ \frac{1 - \delta}{C C_L(1 - L_0) + \sum_{\tau} q[\tau]}, P_C \right\}, \tag{4}
\]

where \( \delta \) is a boolean variable indicating whether the appliance is running at \( \tau = 0 \):

\[
\delta = \begin{cases} 
1 & \text{if the appliance is operating}, \\
0 & \text{if the appliance is not operating}. \end{cases} \tag{5}
\]

Note, that in order for the \( \mu \text{CHP} \) to have ramp-up flexibility (i.e., to increase the power consumption of the household), the appliance must be running and has to be turned off. Vice versa, in order for it to have ramp-down flexibility, the appliance must be idle and has to be turned on. After the buffer is full, the \( \mu \text{CHP} \) can no longer operate. This gives an upper limit to the duration for which the appliance can be running, and thus for the available ramp-down flexibility. When the buffer is drained, the appliance can no longer remain idle, and thereby provides an upper limit to the available ramp-up flexibility.

The same analogy holds with respect to the flexibility of the heat pump. An idle heat pump has to be turned on to provide ramp-up flexibility, whereas a running heat pump has to be turned off to provide ramp-down flexibility. These flexibilities are expressed as:

\[
F^+_{H}[\tau] = \min \left\{ \frac{(1 - \delta)(C_H(1 - L_0) + \sum_{\tau} q[\tau])}{\eta H}, P_H \right\}, \tag{6}
\]

\[
F^+_{H}[\tau] = \max \left\{ -\delta C_H L_0 - \sum_{\tau} q[\tau], -P_H \right\}. \tag{7}
\]

### B. Dual decomposition

We give a brief review on the optimal control problem of a network of \( n \) agents in this section. We first outline the centralized problem, followed by the equivalent distributed formulation. The section is based on [7], [11], and [12].

The network is represented by a weighted, directed graph \( G = (V_n, E_n) \), with \( V_n = \{1, 2, \ldots, n\} \) being the set of agents and \( E_n \subseteq V_n \times V_n \) being the set of edges. The agents are connected in order to exchange information, \( (i, j) \in E_n \) means that agent \( i \) receives information from agent \( j \). The weight on the edge characterizes the importance of the information.

Now consider the discrete-time system

\[
x[k + 1] = Ax[k] + Bu[k] + w[k], \tag{8}
\]

where \( x[k] \in \mathcal{X} \subseteq \mathbb{R}^n \) is the state, \( u[k] \in \mathcal{U} \subseteq \mathbb{R}^n \) is the input, and \( w[k] \in \mathcal{W} \subseteq \mathbb{R}^n \) is the disturbance vector at time-step \( k \). Note, that \( k \) is the general time-step throughout the paper, while \( \tau \) is used to distinguish the number of time-steps it takes to fill the heat buffer from its initial level. \( \mathcal{X}, \mathcal{U}, \) and \( \mathcal{W} \) are bounding sets. Matrix \( A \in \mathbb{R}^{n \times n} \) is the weighted adjacency matrix corresponding to \( G \) with properties

- \( A_{ij} \geq 0 \),
- \( A_{ij} = 0 \), if no information is sent from agent \( j \) to \( i \),
- \( \sum_i A_{ij} = 1 \).

Furthermore, \( G \) is required to be strongly connected. An example is given below, with corresponding topology depicted in Fig. 4.

\[
A = \begin{pmatrix}
0.9 & 0.2 & 0 & 0 & 0 \\
0.1 & 0.6 & 0.1 & 0.2 & 0 \\
0 & 0 & 0.9 & 0.1 & 0 \\
0 & 0.2 & 0 & 0.6 & 0.2 \\
0 & 0 & 0 & 0.1 & 0.8
\end{pmatrix} \tag{9}
\]

Fig. 4. Example information sharing topology between agents. Note, that there is no DSO-like coordinator included in the graph.
We assume that input matrix $B$ is the $n \times n$ identity matrix, $B = I_n$. The objective is to find a sequence of control inputs that minimizes a given quadratic performance index $V(x[k], u[k])$ over $K$ time-steps. We assume this performance index to be separable for the agents, i.e., $V(x[k], u[k]) = \sum_i V_i(x_i[k], u_i[k])$. The problem

\[
\begin{align*}
\text{minimize} & \quad \sum_{k=0}^{K-1} V(x[k], u[k]) \\
\text{subject to} & \quad \text{Eq. (8)}
\end{align*}
\]  

is a centralized problem, in the sense that a central entity collects all relevant information to compute the solution. The main disadvantage of this formulation is that it quickly becomes computationally expensive as we scale the number of agents. Hence, we instead split the problem into smaller sub-problems. The idea is that each agent solves its own sub-problem, based only on local information, and together they arrive to the solution of the original problem. This technique is called dual decomposition.

Let $A_D$ be the matrix of self-weights, $A_D = \text{diag}(a_{11}, a_{22}, \ldots, a_{nn})$, and $A_0 = A - A_D$. We introduce a new variable $v[k]$ so that (8) can be rewritten as

\[
x[k+1] = A_D x[k] + v[k] + u[k] + w[k],
\]  

with the additional constraint

\[
v[k] = A_0 x[k].
\]

We then decompose the problem by applying Lagrangian relaxation, i.e., we augment the objective function with constraint (12), weighted by the vector of Lagrange multipliers $\lambda[k] \in \mathbb{R}^n$. This results in the new objective function

\[
\max_{\lambda} \min_{u,v} \sum_{k=0}^{K-1} V(x[k], u[k]) + \lambda[k]^T (v[k] - A_0 x[k]).
\]

Since $V(x[k], u[k])$ is separable, we can decouple this objective function by interchanging the minimization and summation terms. Thus, each prosumer solves its own separated contribution to the dual-problem

\[
\max_{\lambda_i} \min_{u_i,v_i} \sum_{k=0}^{K-1} \left( V_i(x_i[k], u_i[k]) + \lambda_i[k] v_i[k] - x_i[k] \sum_{j \neq i} A_{ij} \lambda_j[k] \right).
\]

The new formulation is often referred to as the dual problem, with $\lambda[k]$ being the dual variables vector. The solution to the dual problem is the same as the solution to the original problem, if the bounding sets $X, U$, and the objective function $V(x[k], u[k])$ are convex [8]. The Lagrange multiplier is calculated using a subgradient iteration [7], which we will describe in the next section.

IV. SYSTEM DESCRIPTION

A. Device modeling

Here, we describe the models of the $\mu$CHPs and heat pumps used in this paper. The devices have flexible loads ($f_i[k]$), and are modeled as mixed logical dynamical (MLD) systems [21]. Running on fossil fuel, the $\mu$CHP is generating both electricity and heat simultaneously. The amount of electricity produced is assumed to be a constant value $P_C$,

\[
f_i[k] = \begin{cases} 0 & \text{if } \delta_i[k] = 0, \\ P_C & \text{if } \delta_i[k] = 1, \end{cases}
\]

where $\delta_i[k]$ is the generalization of (5) for prosumer (and appliance) $i$ at time-step $k$. The ratio between the produced electricity and heat is $\eta_C$.

When the heat pump is operating, it converts electricity into heat with conversion ratio $\eta_H$. The electricity consumption during the generation process is assumed to be a constant value $P_H$,

\[
f_i[k] = \begin{cases} 0 & \text{if } \delta_i[k] = 0, \\ P_H & \text{if } \delta_i[k] = 1. \end{cases}
\]

In both cases, the generated heat is stored in a heat buffer. The buffer storage level decreases if there is a heat demand in the household, this amount is denoted by $q[k]$. Hence, the storage level can be expressed as

\[
L_i[k+1] = L_i[k] + \frac{\eta}{C} [f_i[k] - \frac{q[k+1]}{C}],
\]

also see Section III-A. The value of $\eta$ is dependent on the type of device.

For efficiency reasons, an appliance is required to keep operating for at least $\tau_{\text{on}}$ time-steps once it had been turned on. Correspondingly, it is required to stay switched off for at least $\tau_{\text{off}}$ time-steps before it can run again. We define counters $t^{\text{on}}_i[k]$ and $t^{\text{off}}_i[k]$ to keep track of the number of time-steps the device is in a given state:

\[
t^{\text{on}}_i[k+1] = \begin{cases} t^{\text{on}}_i[k] + 1 & \text{if } \delta_i[k+1] = 1, \\ 0 & \text{otherwise}, \end{cases}
\]

and

\[
t^{\text{off}}_i[k+1] = \begin{cases} t^{\text{off}}_i[k] + 1 & \text{if } \delta_i[k+1] = 0, \\ 0 & \text{otherwise}. \end{cases}
\]

The bounds on the operating times are thus written as

\[
\tau_{\text{min}} \leq t^{\text{on}}_i[k] \leq \tau_{\text{max}}, \quad \tau_{\text{max}} \leq t^{\text{off}}_i[k] \leq \tau_{\text{max}}.
\]

These dynamical constraints, which contain boolean variables, can be translated into a set of linear inequality constraints, as has been described in [21]. However, with the introduction of the boolean variables, the problem becomes a mixed-integer programming problem, and we lose convexity. The dual gap [8] between the primal and the dual problem is no longer zero. Excess switching between the binary values might happen, therefore we set a maximum number of switches to ensure convergence.

B. Aggregator level

According to the USEF market-based control mechanism, a day-ahead planning (DAP[k]) is made in the plan and validate phases. This is the electricity portfolio the BRP promises to deliver during the operate phase. The actual delivered electricity might deviate from the day-ahead planning due...
to unforeseen events, and the objective of our system is to compensate for this deviation.

First, the day-ahead planning is divided among the aggregators, based on the available flexibility they have. At every time-step, the aggregators receive flexibility information (Section III-A) from their connected prosumers, and accumulate them in order to make the weighting factor:

\[ F_z^+[k] = \sum_{i=1}^{n} F_i^+[k], \quad F_z^-[k] = \sum_{i=1}^{n} F_i^-[k], \]  

where \( i = 1, \ldots, n \) are the prosumers connected to aggregator \( z \). We assume that the number of prosumers are the same for all aggregators. Note, that since we only consider a single appliance per prosumer, we use the same indices for the appliances and the prosumers.

The more flexibility an aggregator has, the bigger share of the day-ahead planning it will receive, so that it contributes more to the balancing problem. Ramp-up flexibility is needed when the electricity demand is lower than the day-ahead planning. In this case, each aggregator receives a day-ahead planning share according to

\[ \text{DAP}_z[k] = \frac{F_z^+[k]}{\sum_i F_i^+[k]} \text{DAP}[k] \quad \text{for all } z. \]  

Similarly, when the electricity demand is higher than the day-ahead planning, ramp-down flexibility is procured. The share of each aggregator is then

\[ \text{DAP}_z[k] = \frac{F_z^-[k]}{\sum_i F_i^-[k]} \text{DAP}[k] \quad \text{for all } z. \]  

The day-ahead planning of each aggregator can be in turn spread among the prosumers based on their available flexibility or evenly. How it was divided did not make much difference in our simulations [22]. For this study, we choose to divide the day-ahead planning evenly.

\[ \text{goal}_i[k] = \frac{\text{DAP}_z[k]}{n} \quad \text{for all } i. \]  

The goal function acts as a reference value for the prosumers during their optimization process.

C. Prosumer level

Here, we describe the optimal control problem within one aggregator. We think of the prosumers as agents, and use the following notation:

- \( f_i[k] \): flexible (controllable) load of prosumer \( i \),
- \( g_i[k] \): fixed (uncontrollable) load of prosumer \( i \).

Electricity load is the sum of supply (production) and demand (consumption), with the convention of using negative sign for supply and positive for demand. The prediction error between the forecasted and actual electricity load is expressed by

\[ \tilde{x}_i[k] = f_i[k] + g_i[k] - \text{goal}_i[k], \]  

or

\[ \tilde{x}_i[k+1] = \tilde{x}_i[k] + u_i[k] + w_i[k] - \Delta \text{goal}_i[k], \]  

where

\[ u_i[k] = f_i[k+1] - f_i[k], \]
\[ w_i[k] = g_i[k+1] - g_i[k], \]
\[ \Delta \text{goal}_i[k] = \text{goal}_i[k+1] - \text{goal}_i[k]. \]  

By introducing the information sharing matrix \( A \), we provide a solution between the prosumers:

\[ x_i[k+1] = A_{ii}x_i[k] + \sum_{j \neq i} A_{ij}x_j[k] + u_i[k] + w_i[k] - \Delta \text{goal}_i[k]. \]  

The requirements on the \( A \) matrix (see Section III-B) ensure that the two are always equal within an aggregator [12],

\[ \sum_{i=1}^{n} x_i[k] = \sum_{i=1}^{n} \tilde{x}_i[k]. \]  

The difference is that \( \tilde{x}_i[k] \) denotes the real, physical imbalance of prosumer \( i \), whereas \( x_i[k] \) is the imbalance information that includes the weighted sum of the neighboring imbalances as well.

1) Distributed optimal control problem: We want to minimize the deviation (error) between the predicted and the actual load, therefore we formulate the (centralized) problem as

\[ \text{minimize} \sum_{i=1}^{K-1} \sum_{k=0}^{n} (\tilde{x}_i[k])^2 \]

subject to

\[ x_i[k+1] = A_{ii}x_i[k] + \sum_{j \neq i} A_{ij}x_j[k] + u_i[k] + w_i[k] - \Delta \text{goal}_i[k], \]  

device-specific constraints from Sec. IV-A.

Note, that the device-specific constraints contain the integer (boolean) variable \( \delta \). In order to obtain the distributed formulation, we decompose the centralized problem by applying (11)-(12):

\[ x_i[k+1] = A_{ii}x_i[k] + v_i[k] + u_i[k] + w_i[k] - \Delta \text{goal}_i[k], \]  

where \( v_i[k] = \sum_{j \neq i} A_{ij}x_j[k] \) is the expected influence of the connected neighbors. The dual objective function that the prosumers have to solve is then given, based on Eq. (14), by

\[ \max_{\lambda_i} \min_{u_i, v_i} \sum_{k=0}^{K-1} \left( x_i^2[k] + \lambda_i[k]v_i[k] - x_i[k] \left[ \sum_{j \neq i} A_{ij}\lambda_j[k] \right] \right). \]  

2) Congestion management: The DSO makes sure that the network is not overloaded. In this paper, we assume that all prosumers (\( N \)) are under one DSO. The total (flexible plus fixed) load of all households should stay within the distribution capacities of the entire network:

\[ \sum_{i=1}^{N} (f_i[k] + g_i[k]) \leq L_{max} \quad \text{for all } k, \]  

where \( L_{max} \) is the maximum network (or transformer) capacity, and is assumed to be constant. The objective function is further augmented with this DSO constraint via Lagrangian relaxation (based on [13]), resulting in
The Lagrange multipliers $\lambda_i[k]$ and $\mu_i[k] \geq 0$ are generally associated with shadow prices [23], and are updated through the subgradient iterations

$$
\lambda_i^{k+1}[k] = \lambda_i^k[k] + \beta_i \left( v_i^k[k] - \sum_{j \neq i} A_{ij} x_i^k[k] \right),
$$

(35)

and

$$
\mu_i^{r+1}[k] = \mu_i^r[k] + \gamma \left( \sum_i (f_i^k[k] + g_i^k[k]) - L_{\text{max}} \right),
$$

(36)

where $r$ is the iteration counter, and $\beta$ and $\gamma$ are appropriately chosen step sizes. In practice, the subgradient iteration is stopped when a sufficiently good approximation of the solution is reached. A stopping criterion is given in [7]; the algorithm terminates if the Lagrangian updates stay within a solution is reached. A stopping criterion is given in [7]; the subgradient iteration. The following cases can happen:

- If there is no congestion detected by the DSO, the Lagrange multipliers have not converged yet, then all aggregators have to perform the optimization loop.
- If there are no congestion points detected, and some, but not all of the $\lambda_i[k]$ values have converged, then only those aggregators have to perform the optimization loop again whose corresponding $\lambda_i[k]$ have not converged.

Note, that in this model the number of prosumers is $n$ for all aggregators, and there is no coupling between prosumers that belong to different aggregators. Using dual decomposition, a multi-level distributed optimization formulation is obtained for the prosumers, who act individually based on local information received from their neighbors. The aggregators only send out the goal functions at the beginning of each optimization loop, they do not act during the process.

### VI. Simulations

#### A. Implementation

We implement our controller in three different scenarios, as explained in Sec. VI-B below. We use MATLAB in combination with Gurobi [25] to solve the mixed-integer quadratic programs (MIQP). A circular topology is considered for the prosumers within an aggregator cluster, with each prosumer having a self-weight of 0.6, and a weight of 0.2 for the information coming from its two neighbors. The corresponding information sharing matrix is

$$
A = \begin{bmatrix}
0.6 & 0.2 & 0 & \ldots & 0 & 0.2 \\
0.2 & 0.6 & 0.2 & \ldots & 0 & 0 \\
0 & 0.2 & 0.6 & \ldots & 0 & 0 \\
0.2 & 0 & 0 & \ldots & 0.2 & 0.6 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \ldots & 0 & 0.2
\end{bmatrix}
$$

(37)

We work with realistic load profiles acquired from pattern generators from the Energy research Centre of the Netherlands (ECN) [26]. Setup parameters for the appliances are derived
Fig. 5. (Scenario 1) Simulation 1, without congestion management. The upper figure illustrates the total load within the prosumer network, the lower figure depicts the flexible and fixed loads in detail. The system aims to follow the forecasted day-ahead planning, the mismatch is shown as the green line. The network capacity limit is violated several times.

Fig. 6. (Scenario 1) Simulation 2, with congestion management. All of the DSO constraint violations are resolved by shifting the production and consumption of the flexible devices in time.

from the devices installed in PowerMatching City, a smart grid demonstration project in the Netherlands [27]. Table I summarizes these parameters. We initialize all Lagrangian multipliers with zero, ε, the subgradient convergence criterion for both multipliers, is chosen to be 0.04.

T_{\text{on}}^{\text{max}} and T_{\text{off}}^{\text{max}} are determined by the physical limits of the heat storage buffers attached to the devices: if the heat buffer is empty, the appliance has to turn on, and if the buffer is full, the appliance has to stop operating. In the simulation, one time-step corresponds to 15 minutes, and the prediction horizon for the d-MPC is taken to be 8 time-steps. The buffer levels are arbitrarily initialized. The non-summable diminishing step sizes [8] of the subgradient iterations are chosen to be equal, \( \beta_i^r = \gamma^r = \mu / \sqrt{r} \). Different \( \gamma_0 \) choices have been tested in [28], for our simulations, we have chosen \( \gamma_0 = 1 \).

For demonstration purposes, we set the network capacity limit to a value that is lower than the standard level (i.e., an approximate 1.1 kW per household). This is to better see the effects the flexible appliances can have in resolving congestion within the distribution network. In certain cases in our demonstration simulations, the DSO constraint is lower than the day-ahead planning. In these situations, the day-ahead planning is shaved because of the USEF requirement, that the DAP has to be valid (i.e., congestion free).

B. Results

1) Scenario 1: In this scenario, the network consists of 6 households, each with one appliance: 3 are equipped with microCHPs, 3 with heat pumps, and are connected in an alternating order.
In Simulation 1 (Fig. 5), the DSO constraint is ignored. The figure depicts the total electricity load of the network over the simulation period, as well as the individual loads of the prosumers. Note, that the local production from the \( \mu \)CHPs lowers the total load. The network capacity limitation is violated at time-steps 10-12, 24, and 47-48. The difference between the forecasted and the actual load (over the whole simulation period) is kept as low as possible.

In Simulation 2, we apply congestion management to the system (Fig. 6). Three different ways of resolving congestion are observed:

- In time-steps 10-11, the \( \mu \)CHP of Prosumer #2 is shifted backwards in time (from time-steps 13-15), in order to lower the high fixed load. Only two of the initial three instances of operation can be shifted due to its buffer capacity limits (Fig. 7). Congestions caused by fixed electricity loads can only be resolved by turning on \( \mu \)CHPs. In time-step 24, the two \( \mu \)CHPs were not able to bring the total load under the network capacity limit, therefore the third \( \mu \)CHP is kept operating at that time-step. The total production of the three \( \mu \)CHPs together is enough to resolve the congestion.
- In time-step 12, the consumption of the heat pump of Prosumer #5 is moved away to time-steps 6-7, in order to avoid congestion.
- In time-steps 47-48, the heat pumps cannot be moved away because of the high heat demands of the respective prosumers, but the congestion can be resolved by \( \mu \)CHPs. The electricity production of the \( \mu \)CHPs compensate for the consumption of the heat pumps.

Congestion can be resolved by moving \( \mu \)CHPs to the time of violation, or by moving heat pumps away from those instances.

If we look at the buffer levels (Fig. 7), we notice that because the \( \mu \)CHP of Prosumer #2 is only running for two time-steps in Simulation 2 instead of three, the heat demand drains its buffer enough for the device to be able to operate for a longer duration from time-step 20 onwards. This helps to resolve the network capacity violation at time-step 24. Another observation is that at time-step 33, the heat buffer of Prosumer #3 is turned on, although the day-ahead planning is low. This is due to the fact that its heat buffer is empty, therefore the heat pump must turn on regardless of the predicted load. Moreover, notice that in Simulation 2 (with DSO constraint), Prosumer #2 makes more use of the maximum storage capacity of the heat buffer, thereby providing more ramp-down flexibility to resolve congestion.

Fig. 8 shows the evolution of \( \lambda_1, ..., \lambda_6 \), and \( \mu \) at time-step 8, when congestion is detected within the prediction horizon of the d-MPC. The shadow price \( \mu \) keeps oscillating between updating positively when a congestion is detected, and updating negatively when there is no congestion detected until the \( \lambda_i \)s are converged. It ensures that all congestion in the prediction horizon is prevented, and indirectly reflects the price the DSO has to pay to the prosumers during the USEF settlement phase, to compensate for their modified behavior.

2) Scenario 2: In the second scenario, we run the simulations with 1,000 prosumers to show that the algorithm is capable to handle larger networks.

In Simulation 1 (Fig. 9), the DSO constraint is violated, which is then solved in Simulation 2 (Fig. 10). We can observe that at those time-steps where there is congestion in Simulation 1, more \( \mu \)CHP devices are turned on in Simulation 2 to decrease the total load of the network, and hence resolving the congestion.

The running time of the simulation does not scale linearly with the number of prosumers. This is due to the fact that although the algorithm is distributed, the simulation itself is not implemented in a parallel fashion. In reality, we envision that
each prosumer will have a smart meter or computer in their homes, and the calculations can be done in parallel. Only the DSO has to collect the information about the prosumer loads, the prosumers only work with local information from their neighbors. Therefore, that part of the algorithm is scalable.

3) Scenario 3: Scenario 3 presents the hierarchical control case, where multiple aggregators are involved. In this scenario, 2 aggregators, with 50 prosumers each, are considered. The distribution line capacity limitation is again violated in Simulation 1 (Fig. 11). In Simulation 2 (Fig. 12), the algorithm solves the congestion points by proportionally distributing the day-ahead planning among the two aggregators, based on their available flexibility. Both aggregators (and their prosumers) contribute to congestion management, which can be seen from the sub-division of the prediction error between the aggregators: in Simulation 1, the total prediction error (the objective function) is 389.98 kW² and 372.65 kW², respectively, on the two aggregators. In Simulation 2, the total prediction error on each aggregator is 408.21 kW² and 411.11 kW². In the latter simulation, the additional constraint on the distribution line capacities makes the problem tighter, and thus increases the value of the objective function.

Table II summarizes the maximum number of convergence iterations.
iterations for each scenario. It is notable that the second simulations (with congestion management) take a higher number of iterations to converge. This is expected, as the limitations on the distribution lines add another constraint to the problem, making it more difficult to solve.

VII. CONCLUSION

We formulate an optimal control problem with the goal of minimizing the error of prediction for supply and demand in a USEF-compliant network. In its hierarchical structure, we quantify flexibility in order to divide the day-ahead planning among the aggregators, and then develop a distributed model predictive controller to achieve our objective. The prosumers share local information with each other, thus reach the solution in a cooperative manner. The goal function, derived from the day-ahead planning, provides the link between the different phases of the USEF market-based mechanism. We demonstrate in various simulation scenarios that our model successfully avoids congestion by procuring flexibility from the smart appliances, while keeping the prediction error to a minimum.

There is ongoing work to expand the scope of the model and add optimization to the upper levels (BRP and aggregator) as well, thereby obtaining a multi-level optimal control problem. We aim to take different objectives on the different levels, i.e., the stakeholders can optimize towards their own interests. One possible way in this direction is to incorporate a pricing mechanism.

VIII. ACKNOWLEDGMENTS

The authors thank Wouter Kramer for his help in preparing this manuscript, and Gunn K. H. Larsen for her contribution to the preliminary results.