

A Distributed Hierarchical Architecture for Community-based Power Balancing

Rodrigo Verschae, Hiroaki Kawashima, Takekazu Kato and Takashi Matsuyama

Graduate School of Informatics

Kyoto University, Kyoto, 6068501, Japan

Emails: {rodrigo,tkato}@vision.kuee.kyoto-u.ac.jp, {kawashima,tm}@i.kyoto-u.ac.jp

Abstract—The coordination and control of the agents in electrical networks is expected to have an important positive impact in future smart-grids. In the present paper we propose a distributed hierarchical coordination framework for power balancing, and the algorithms required to build such coordination architecture. This architecture is built such that a group of agents (a subtree of the hierarchy) can respond and adjust to deviations from a day-ahead plan for that group, without the need to coordinate with its upper levels of the hierarchy at every time step. The agents are arranged in groups by taking into account a measure of the flexibility of each group, thus allowing a partial de-coupling of the groups during the coordination, which reduces the required work of the central coordinator and gives robustness to the system. The assignment of agents to groups is formulated as a welfare maximization problem, a NP-hard problem for which an approximate solution can be obtained efficiently by using the sub-modularity and monotonicity of the involved functions. An electric vehicle charging coordination scenario is presented to exemplify the proposed architecture and methods.

I. INTRODUCTION

The balancing of generation and consumption patterns is crucial for the operation of the electric power system (e.g. in the frequency stability of the network), and for reducing the generation costs. To perform such balancing, the traditional solution has been to modulate the generation, and to a smaller degree, to use demand response methods and dynamic pricing schemes [1][2][3]. The introduction of smart-grid technologies [4]—in particular the ability to sense in real-time the power usage and generation of every network agent in a energy management system (appliances, storage, etc.)—, together with the growing capability for controlling and scheduling the power usage of those agents, and the increasing use of renewable local generation and mobile loads, is allowing and requiring balancing techniques that can be managed by the demand side instead of being managed by the utility. This, in addition to open new market opportunities [5], and together with the rapid introduction of renewable local generation, local storage, and electric vehicles (EV), is pushing to completely rethink and redesign the role of the stakeholders in power networks [4].

We are interested in allowing demand-side coordination and control of power consumption and generation, where a group of agents (households, factories, dwellings, etc.) can coordinate and balance its aggregated power consumption patterns. Thus, allowing the group of agents to reduce their energy consumption cost¹ and environmental impact, and also

to increase the predictability of their aggregated power usage, with all the benefits that this brings to the grid.

Within this context we propose a distributed hierarchical coordination framework for community power balancing that builds on [6] (briefly presented in Section II), where a distributed coordination profile-based negotiation was introduced. We start by noting (in Section III) that the coordination problem in [6] can be reformulated as a nested *sharing problem*, where the coordination can be done among groups of agents, and where each of these groups solves a smaller sharing problem. This allows each group of agents to partially adjust to deviations from its planned power usage profile, without the need to coordinate with other groups at every time step. The hierarchical architecture is built (in Section IV) by grouping the agents seeking that the *flexibility* of each group is maximized, so that each group can adjust to deviations from their planned power profile. This is formulated as a welfare maximization problem [7], for which an approximated solution can be obtained efficiently using the sub-modularity [7] [8] of the involved functions. For doing the grouping, a measure of the *flexibility* of a group, based on the entropy of the distribution of its aggregate profile, is proposed. To exemplify the proposed architecture, we present an EV coordination scenario (in Section V) where a group can deal with deviations without the need to coordinate with upper levels of the hierarchy at every time step. In Section VI we conclude and discuss possible extensions, such as complementary welfare measures to use during the grouping in order to take into account the topology of the distribution network and the geographical dispersion of the agents.

We assume an incentive-based program where a group of participants get a reward for achieving some demand-shape profiles, or where the group of participants tries to minimize its aggregated power usage cost in a price-based demand response program. Thus, a group of users (a community) coordinate its aggregated power consumption profile for a given period of time (e.g. for the next day) and commits to follow that profile. The coordination is done by taking into account a global cost function associated to the aggregated profile, but also the consumption patterns and preferences of each user. We assume that the appliances coordinate in advance their power usage (e.g. formulating a day-ahead plan), but we also assume that some users may not be able to achieve their planned profiles, because of changes in usage pattern, failures or emergencies. In this case the original plan can not be followed, and thus with help of the remaining users, the community must coordinate to try to remain as close as possible to its original plan.

¹We assume the community buys energy from the utility, and that this cost is related to its power usage pattern. We do not analyze how this cost is calculated or split, but we focus on how to coordinate the power usage pattern.

II. PRELIMINARIES

In [6], a distributed architecture for coordinating the power usage of a community of agents was introduced. Two main ideas were proposed: *i)* the use of power profile based distributed coordination, and *ii)* the modeling of Quality of Life (QoL) of a household using a generative probabilistic model of the power consumption of each agent. The coordination was formulated as an optimization problem that takes into account the local cost of each agent to achieve a given power, and a global cost that measures the power balance of the community. This is illustrated in Fig. 1, where a group of agents $i \in \mathcal{N} = \{1, \dots, N\}$, together with a coordinator, iteratively negotiate their future aggregated power usage. In the following we give a brief overview of the framework and methods, and direct the reader to [6] for further details.

Distributed Coordination: The coordination done by the community is formulated as a *sharing problem*:

$$\underset{(x_i)_{i \in \mathcal{N}}}{\text{minimize}} \left[\sum_{i \in \mathcal{N}} f_i(x_i) + g_r\left(\sum_{i \in \mathcal{N}} x_i\right) \right]. \quad (\text{P1})$$

The agents $i = 1, \dots, N$ in the community coordinate their power usage, each with a power profile $x_i \in \mathbb{R}^T$, and where T is the of number time slots (e.g. 10 min time slots and $T = 144$ represent a 24 hour period). The function $f_i : \mathbb{R}^T \rightarrow \mathbb{R}$ measures the local cost (or difficulty) of agent i to achieve the profile x_i , while the function $g_r : \mathbb{R}^T \rightarrow \mathbb{R}$ measures the cost for the community to have an aggregated profile $\sum_i x_i$.

For scalability and privacy issues, it is desirable that the function f_i is only available to agent i , while g_r is only available to the coordinator, thus a distributed coordination is used. By introducing duplicate variables $z_i = x_i \forall i \in \mathcal{N}$, and under some assumptions for the functions f_i and g_r (e.g. convexity), the Problem (P1) can be solved iteratively using ADMM [9], where at iteration k the following is calculated:

$$\begin{aligned} x_i^{k+1} &:= \arg \min_{x_i} \left[f_i(x_i) + \frac{\rho}{2} \|x_i - x_i^k + b^k\|_2^2 \right], \forall i \in \mathcal{N} \\ \bar{z}^{k+1} &:= \arg \min_{\bar{z}} \left[g_r(N\bar{z}) + \frac{N\rho}{2} \|\bar{z} - \bar{x}^{k+1} - \nu^k\|_2^2 \right] \\ \nu^{k+1} &:= \nu^k + \bar{x}^{k+1} - \bar{z}^{k+1}, \end{aligned} \quad (1)$$

where for a set $\{a_i\}_{i \in \mathcal{N}}$, we note \bar{a} to represent its average (i.e. $\bar{a} = \frac{1}{N} \sum_{i \in \mathcal{N}} a_i$). The vector $\nu^k \in \mathbb{R}^T$ corresponds to the scaled Lagrange multipliers [9], ρ is a parameter, and $b^k = \bar{x}^k - \bar{z}^k + \nu^k$ is a broadcast signal. The first step (x_i^{k+1} -step) is evaluated concurrently by all agents (agent i only needs to know b^k), while the second and third steps are evaluated by the coordinator, which only needs to aggregate $(x_i^{k+1})_i$, calculate \bar{x}^{k+1} , \bar{z}^{k+1} and ν^{k+1} , and then broadcast b^k to all agents.

Local Cost: Each agent i has a local cost of the form:

$$f_i(x_i) = \min_{u_i} \left[f_i^u(u_i) + f_i^{x|u}(x_i, u_i) \right], \quad (2)$$

which is derived from a generative probabilistic model $P(x_i, u_i) = P(u_i)P(x_i|u_i)$. The probability $P(u_i)$ measures how natural control $u_i \in \mathcal{U}_i$ is, while $P(x_i|u_i)$ is a measure of the certainty of achieving the profile x_i given the control signal u_i . Thus, the local cost is defined as:

$$f_i(x_i) = \min_{u_i \in \mathcal{U}_i} [-\log P(x_i, u_i)], \quad (3)$$

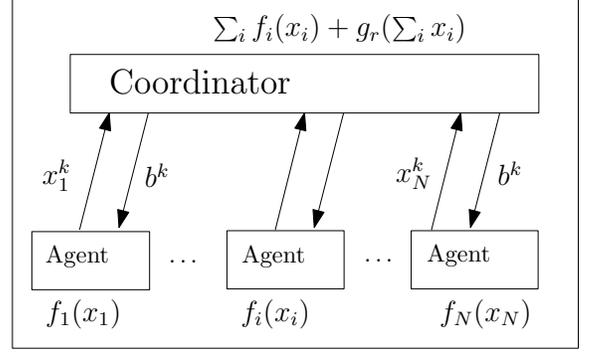


Fig. 1. Two-level distributed coordination architecture proposed in [6].

with $f_i^{x|u}(x_i, u_i) = -\log P(x_i|u_i)$, and $f_i^u(u_i) = -\log P(u_i)$. In case $P(x_i, u_i)$ becomes zero, $f_i(x_i)$ becomes $+\infty$, indicating that the associated profile x_i is not achievable. An example of this is when $P(x_i|u_i)$ is given by the delta function $\delta(x_i - \chi_i(u_i))$, with $\chi_i : \mathcal{U}_i \rightarrow \mathbb{R}^T$ some mapping from control signals u_i to profiles x_i .

Given that two key aspects for the balancing of power profiles are to schedule and to control the load, the probability of the control u_i , $P(u_i)$, is modeled using a hidden semi-Markov model (HSMM) [10] that explicitly models time-varying signals as a sequence of time intervals (or modes). Thus, the profile x_i is described by a sequence of discrete states $s_{i,t} \in \mathcal{Q}_i = \{q_{i,m}\}_{m=1, \dots, M_i}$, where each state represents the state (or mode) of appliance i at time t . We assume that the control variables u_i are these modes at time $t = 1, \dots, T$. In [6] it is described how to efficiently solve Eq. (1.1) for f_i as defined in Eq. (3) using HSMM segment models.

Global Cost: The global cost function, $g_r(v)$, measures the cost associated to the aggregated power usage $v \in \mathbb{R}^T$ with respect to a reference profile $r \in \mathbb{R}^T$. The profile r can represent the expected generated power profile or the planned consumption profile, and when $r = 0$ the goal becomes to flatten the power consumption (when all agents are consumers). Two examples of such functions are:

- the peak-to-average ratio (PAR): $g(v) = \alpha \frac{T \|v\|_\infty}{\sum_t v_t}$, and
- the Euclidean squared norm of the deviation from a reference profile: $g_r(v) = \alpha \|v - r\|^2$.

III. HIERARCHICAL DISTRIBUTED COORDINATION

We use a hierarchical distributed coordination to allow a partial de-coupling of non-overlapping subtrees of the hierarchical structure. For simplicity we will present the hierarchical formulation for a two-level coordination (three-level architecture), but extending it to deeper hierarchies is straight forward.

Formulation: Let us assume there are J groups of agents, and let $\mathcal{J} = \{1, \dots, J\}$ be the index set representing these groups. (**Notation:** In the following we use the index j to refer to groups and the index i to agents). Let $\{\mathcal{N}_j\}_{j \in \mathcal{J}}$ be a partition of \mathcal{N} (i.e. $\mathcal{N} = \cup_{j=1}^J \mathcal{N}_j$ and $\mathcal{N}_j \cap \mathcal{N}_k = \emptyset$ for $j \neq k$), with $N_j = |\mathcal{N}_j|$ the cardinality of group j and $N = \sum_{j=1}^J N_j$ the total number of agents. Considering the partition $\{\mathcal{N}_j\}_{j \in \mathcal{J}}$,

now we rewrite problem (P1) as:

$$\begin{aligned} \text{minimize} \quad & \sum_{(x_i)_{i \in \mathcal{N}}} \sum_{j \in \mathcal{J}} \sum_{i \in \mathcal{N}_j} f_i(x_i) + g_r \left(\sum_{j \in \mathcal{J}} \sum_{i \in \mathcal{N}_j} x_i \right) \\ & + \sum_{j \in \mathcal{J}} \hat{g}_j \left(\sum_{i \in \mathcal{N}_j} x_i \right), \end{aligned} \quad (\text{P2})$$

where, in addition to the cost of each agent (represented by functions $f_i(x_i)$), and the global shared cost $g_r(\sum_{j \in \mathcal{J}} \sum_{i \in \mathcal{N}_j} x_i)$, we have introduced an extra cost term $\hat{g}_j(\sum_{i \in \mathcal{N}_j} x_i)$ for each group $j \in \mathcal{J}$. Each of these extra terms can encode a shared cost for the group, a set of constraints for the group (e.g. $\hat{g}_j(x) = 0$ for $x \in \mathcal{X}_j$, and $+\infty$ otherwise), or could be zero for all x , which corresponds to problem (P1).

Let us now introduce the variables w_j which corresponds to the duplicate of the aggregated profile of group j , $\sum_{i \in \mathcal{N}_j} x_i$, which together with the corresponding required constraint leads to the equivalent optimization problem:

$$\begin{aligned} \text{minimize} \quad & \sum_{(x_i)_{i \in \mathcal{N}}} \sum_{j \in \mathcal{J}} \left[\sum_{i \in \mathcal{N}_j} f_i(x_i) + \hat{g}_j \left(\sum_{i \in \mathcal{N}_j} x_i \right) \right] + g_r \left(\sum_{j \in \mathcal{J}} w_j \right) \\ \text{s.t.} \quad & w_j - \sum_{i \in \mathcal{N}_j} x_i = 0, \quad \forall j \in \mathcal{J}. \end{aligned} \quad (\text{P3})$$

It is not difficult to observe that this corresponds to a sharing problem where $g_r(\sum_{j \in \mathcal{J}} w_j)$ is the cost shared among J groups, and that the coordination can be performed among groups of agents. (**Notation:** the subscript g will be used to refer to variables that are calculated at the global coordinator (with the exception of w_j^g), while a hat ($\hat{\cdot}$) or a tilde ($\tilde{\cdot}$) for some of the variables calculated at each group coordinator). When applying the ADMM method, we obtain the following iterative procedure:

$$\begin{aligned} (\tilde{x}_i^{q+1})_{i \in \mathcal{N}_j} &:= \arg \min_{(x_i)_{i \in \mathcal{N}_j}} \left[\sum_{i \in \mathcal{N}_j} f_i(x_i) + g_j^q \left(\sum_{i \in \mathcal{N}_j} x_i \right) \right], \quad \forall j \in \mathcal{J} \\ \bar{w}_g^{q+1} &:= \arg \min_{\bar{w}_g} \left[g_r(J\bar{w}_g) + \frac{J\rho_g}{2} \|\bar{w}_g - \bar{x}^{q+1} - \nu_g^q\|_2^2 \right] \\ \nu_g^{q+1} &:= \nu_g^q + \bar{x}^{q+1} - \bar{w}_g^{q+1}, \end{aligned} \quad (4)$$

with q the iteration index at the upper level. Here $\bar{w}_g^q = \frac{1}{J} \sum_{j \in \mathcal{J}} w_j^q$, and \tilde{x}_i^q is the profile of agent i at iteration q (after optimization problem in group j has finished). The remaining variables are defined as $\bar{x}^q = \frac{1}{J} \sum_{j \in \mathcal{J}} \hat{x}_j^q$, with $\hat{x}_j^q = \sum_{i \in \mathcal{N}_j} \tilde{x}_i^q$, and ρ_g the global penalty of the quadratic term. It is important to note that the problem solved in the (\tilde{x}_i^{q+1}) -step of this iterative procedure also corresponds to a sharing problem, with shared cost:

$$g_j^q \left(\sum_{i \in \mathcal{N}_j} x_i \right) = \hat{g}_j \left(\sum_{i \in \mathcal{N}_j} x_i \right) + \frac{\rho_g}{2} \left\| \sum_{i \in \mathcal{N}_j} x_i - \hat{x}_j^q + b_j^q \right\|_2^2, \quad (5)$$

where $b_j^q = \bar{x}^q - \bar{w}_g^q + \nu_g^q$. That is, the shared cost of group j consists of the original cost of the group plus a quadratic penalty related to the deviation from the reference of the upper level cost. Thus, we have decomposed the original sharing problem into a sharing problem solved among groups plus a set of smaller sharing problems solved by the groups.

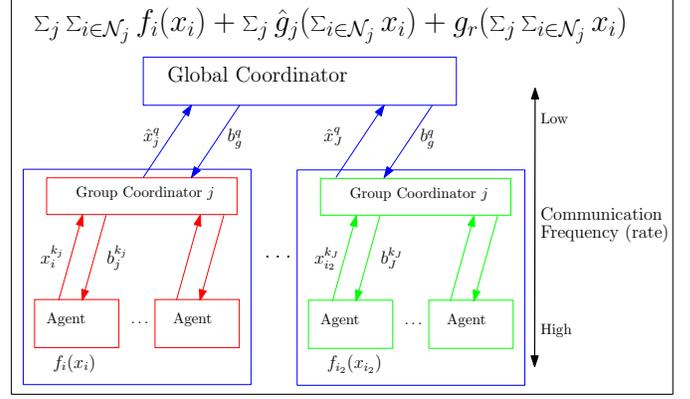


Fig. 2. Hierarchical distributed architecture.

The groups collaborate to solve a sharing problem in a distributed way, while the agents in a group collaborate to solve a sharing problem for the group. This naturally translates in a nested architecture as presented in Fig. 2. In the figure, different colors indicate different sharing problems being solved and their relations: blue (outer boxes) indicate the sharing problem solved by the groups, while red and green (inner boxes) correspond to smaller sharing problems solved within groups. Two advantages of this nested architecture are that: *i*) the problems solved within two different groups are decoupled, which implies that agents from different groups do not need to work synchronously, and *ii*) the upper levels can work at a lower frequency than the lower levels, thus reducing the load on the global coordinator (as indicated in Fig. 2).

Coordination within each group: The optimization problem within each group in Eq. (4.1) considers a group cost that includes the quadratic term coming from the upper-level. For group j , the algorithm obtained using ADMM (with $b_j^k = \bar{x}_j^k - \bar{z}_j^k + \nu_j^k$) is:

$$\begin{aligned} x_i^{k+1} &:= \arg \min_{x_i} \left[f_i(x_i) + \frac{\rho_j}{2} \|x_i - x_i^k + b_j^k\|_2^2 \right], \quad \forall i \in \mathcal{N}_j \\ \bar{z}_j^{k+1} &:= \arg \min_{\bar{z}_j} \left[g_j^q(N_j \bar{z}_j) + \frac{N_j \rho_j}{2} \|\bar{z}_j - \bar{x}_j^{k+1} - \nu_j^k\|_2^2 \right] \\ \nu_j^{k+1} &:= \nu_j^k + \bar{x}_j^{k+1} - \bar{z}_j^{k+1}, \end{aligned} \quad (6)$$

where the minimization solved at the z -step in Eq. (6) is the sum of three terms (see also Eq. (5)), including two quadratic terms: the deviation from the group's goal, and the deviation from the global goal. Note that index q is used for the iteration of ADMM at the upper, group level. We write the same k for the iteration index within the groups, but when reading k it should be understood as k_j , the current iteration within group j (the iterations within a group are not synchronized with the iterations within other groups).

Since in the present work we are more interested in the hierarchical architecture than in the structure of the cost functions, in the following we consider that there is no cost for the group, i.e. $\hat{g}_j(\cdot) = 0$. In this case Eq. (6.2) can be solved analytically:

$$\bar{z}_j^{k+1} = \frac{1}{1 + \alpha_{\rho_j}} \left[\alpha_{\rho_j} (\bar{x}_j^{k+1} + \nu_j^k) - r_j^q \right], \quad (7)$$

with $\alpha_{\rho,j} = \frac{\rho_j}{\rho_g N_j}$, and $r_j^q = \frac{1}{N_j}(b_g^q - \hat{x}_j^q)$. In other words, \bar{z}_j^k and ν_j^k are linear in all the values required to be evaluated (including the residuals using in the algorithm's stopping conditions [9]). Thus, when there is no cost for the group, the group coordination is simpler than the global one, because the state of the negotiation can be easily calculated by any entity that has access to the sequence $(\hat{x}_j^k)_{k=0,\dots,K_j}$ and to r_j^q , with K_j the number of iterations. Note that r_j^q has the role of being the global reference profile for group j at iteration q .

On-line coordination in the hierarchical architecture: So far we have described how the coordination can be solved in a nested fashion using the hierarchical architecture, where groups of agents negotiate their power usage, and within each group a coordination is done among the agents in the group. Now we modify the scenario and assume that all groups (and all agents in the groups) have already negotiated their power usage profiles, and that they need to adjust when one or more agents deviate from their planned power profiles. We assume that each group has a planned power reference profile $r_j = \sum_{i \in \mathcal{N}_j} r_i$ that was determined in advance, e.g. a day-ahead schedule (e.g. $r_j = \hat{x}_j^*$, with \hat{x}_j^* the solution of the last iteration of ADMM solved using Eq. (4)). We will call $r = \sum_{j \in \mathcal{J}} r_j$, the reference profile for the community, and we assume the reference profiles are fixed over time during the period $[1, T]$.

We want each group to be partially decoupled from the global coordination, such that when an agent $i \in \mathcal{N}_j$ changes its plan, the group j can adjust for a time frame of duration Δ , before the group performs a global adjustment with other groups. In cases when the deviation is small, the group is expected to be flexible enough to completely absorb this deviation, such that no change is observed by other groups, the global coordinator or the utility, while when the deviation is large, the group will absorb part of the deviation, and later all groups will help to respond to the change. Thus, a subtree of the hierarchical architecture should be able to handle some unplanned changes. There are two main goals for implementing such behavior:

- To reduce the communication and processing load of the global coordinator. This would also allow higher communication rates within the group – when compared with the groups coordination in the upper level. As an extreme example, consider the case when there is a global coordinator and no groups (as in [6]): this would put a rather high load in the coordinator when the number of agents is large, and the dependency of the whole system on the global coordinator would be too high.
- To give robustness to the system. In case of communication problems among one group and the global coordinator (e.g. during emergencies, failures, etc.), that group (a subtree of the hierarchical architecture) would still be able to coordinate and respond to deviations within the group. Thus, the coordination within a group allows its agents to help stay close to their planned reference profile, and therefore the system would also stay close to its global reference, even if there are communication issues between a group and the global coordinator.

The described on-line coordination assumes that: *i*) a group is flexible enough to adjust to not-too-large deviations within the group, and *ii*) the agents in a group can communicate

within the group at a fast enough rate (with low delay). This last point may require some geographical closeness for the agents within a group, which is also required to better respond in case of failures in the communication network. Also, and even though not strictly achievable unless cost functions are assigned to each group, if some agents in a group receive power through the same branch of the distribution network tree (e.g. they both depend on a subset of substations), the coordination among the agents in the group would reduce the chance to overload those nodes (substations) in the distribution network (with a goal similar to [11]).

In the following section we present a general method for grouping the agents, and give an explicit implementation that takes into account the flexibility of the groups. Nevertheless the framework could also include other goals, such as geographical closeness and path sharing in the distribution network.

IV. BUILDING OF THE HIERARCHICAL ARCHITECTURE

Flexibility of a group: We understand flexibility² as the capacity to absorb fluctuations, and the ability to cope with changes and to accommodate uncertainty. In general, this capacity is related to the diversity of a system, and in its ability to respond to change.

In our case, we want each group to be able to adjust when one or more agents in the group deviate from the planned profile. This can be seen as the ability of each group to have various power profiles, or in other words, to be able to generate an aggregated power profile with a large variability of shapes. To measure this capacity, we consider the distribution of the aggregated profile that can be produced by a group, and we seek to have groups with a distribution as “spread” as possible. We therefore need a measure of the spread of such distribution, and for this we propose to use the *entropy* of the group's aggregated profile as a measure of the flexibility of the group.

Formally, let us call $X_i \in \mathbb{R}^T$ the multivariate random variable associated with the profile of agent i , and $Y_j = \sum_{i \in \mathcal{N}_j} X_i$ the random variable associated to the aggregated profile of the agents in group j . We want the entropy of the aggregated profile, $H(Y_j)$, to be large for every group $j = 1, \dots, J$, so that all groups are flexible enough.

Partitioning procedure (welfare maximization problem): For building the hierarchical architecture we consider a top-down procedure. We assume that we have a set of agents, \mathcal{N} , to be partitioned into J groups. Thus, recursively applying this procedure to the obtained groups we can build the hierarchical structure. For simplicity we present the procedure for building a three-level architecture, but building deeper trees is straight forward. This framework can accommodate not only the flexibility measure proposed above, but also other measures. In the following we present the procedure for the general case.

We want to partition the agents in $\mathcal{N} = \{1, \dots, N\}$ into J groups. This is formulated as the following problem:

$$\arg \max_{(\mathcal{N}_1, \dots, \mathcal{N}_J) \in \mathcal{P}(\mathcal{N})} \left[\sum_{j=1}^J W_j(\mathcal{N}_j) \right], \quad (\text{P4})$$

²The flexibility in production systems has been widely studied in the field of manufacturing [12][13], where the ability to adjust to changes in relevant factors (product, process, loads and failure) is of productive importance.

which is known as the welfare maximization problem, where we seek to maximize the sum of the welfare of the different groups over all possible assignments $(\mathcal{N}_1, \dots, \mathcal{N}_j)$, i.e. over all possible elements of the partition set of \mathcal{N} , $\mathcal{P}(\mathcal{N})$. The set function $W_j : 2^{\mathcal{N}} \rightarrow \mathbb{R}_+$ evaluated in a set $S \subseteq \mathcal{N}$ gives the welfare for group j when being assigned the agents in S .

Solving the assignment problem in (P4) is NP-hard [7], and thus an approximate algorithm must be used when the number of agents is large, which is our case. However, efficient approximate algorithms for the welfare problem exist when the functions W_j are submodular³ and monotone, and in such case the problem (P4) corresponds to a problem known as the submodular welfare maximization [7], for which approximation algorithms exist in the off-line [8] and on-line setting [7]. The existing approximated algorithms are γ -approximations, which means that in expectation the solution they give is at least γ OPT, with OPT the welfare for the optimal assignment. In the current work we consider the off-line setting, for which there is an optimal approximation that has the best achievable bound in the off-line setting [8] and has $\gamma = (1 - 1/e) \approx 0.632$ (in the on-line setting $\gamma = 0.5$ [7]).

In the off-line setting, the optimal approximation is obtained (in expectation) by allocating each agent in groups at random following a distribution estimated using the welfare functions $W_j(S) \forall j$ [8]. In the case the set function is the same for all groups, $W(S) = W_j(S) \forall j$ (as in the case of the entropy), the optimal approximation is obtained by allocating the agents uniformly at random. To obtain a good solution, this random assignment of agents to groups is repeatedly applied, and the best observed assignment is kept. Following [8] we call this allocation procedure *uniform assignment algorithm*.

The entropy of a sum of multivariate random variables, defined in the previous subsection to measure the flexibility of a group of agents, evaluates the *entropy of sums*, which is known to be submodular and monotone [14], and thus can be used as welfare measure within the presented framework. (Note that this is not the same as the entropy of a joint probability distribution, which is more common and also known to be submodular [15]). In our current setting, the welfare function is designed to measure the flexibility of a group, but other measures are also possible. In addition, it is important to mention that any nonnegative linear combination of submodular (monotone) functions gives a nonnegative submodular (monotone) function [7], and thus a more general welfare measure can be obtained by a nonnegative linear combination of such functions.

Log determinant of the covariance matrix: Now we analyze the entropy as flexibility measure for a particular case. Let us assume that all X_i , with $i = 1, \dots, N$, are multivariate normal (MVN) random variables. Then, the aggregated profile Y_j is also a MVN, and its entropy is proportional to the log determinant of its covariance matrix of Y_j : $H(Y_j) \propto \log |\mathcal{K}_{Y_j}|$, up to an irrelevant constant additive factor that depends on the dimension T . Thus, in this case we can define the function set $w : 2^{\mathcal{N}} \rightarrow \mathbb{R}_+$ measuring the flexibility of a group \mathcal{N}_j as:

$$W(\mathcal{N}_j) = \log |\mathcal{K}_{Y_j}|, \quad (8)$$

³A set function $f : 2^{\mathcal{N}} \rightarrow \mathbb{R}$ is *monotone* if $f(S) \leq f(P)$ whenever $S \subseteq P$. We say that f is *submodular*, if for all S_1, S_2 and s such that $S_1 \subseteq S_2$ and all $s \notin S_2$ we have that $f(S_1 \cup s) - f(S_1) \geq f(S_2 \cup s) - f(S_2)$.

with $Y_j = \sum_{i \in \mathcal{N}_j} X_i$, and $X_i \in \mathbb{R}^T$ the profile of agent i .

Under the assumption that the agents are independent MVN random variables, the covariance matrix of the aggregated profile of a group \mathcal{N}_j can be expressed as the sum of the covariance matrix of the agents in that group: $\mathcal{K}_{Y_j} = \sum_{i \in \mathcal{N}_j} \mathcal{K}_i$, with \mathcal{K}_i the covariance matrix of agent i . Thus, the flexibility of a group becomes: $W(\mathcal{N}_j) = \log \left| \sum_{i \in \mathcal{N}_j} \mathcal{K}_i \right|$.

There are several ways to estimate of the covariance matrices $\{\mathcal{K}_i\}_i$. For example, every agent i could estimate its matrix \mathcal{K}_i and send it to the coordinator. Another option, which is used in the simulations in the following section, corresponds to use the history of profiles obtained during the day-ahead negotiation $\{x_i^k\}_{k=1, \dots, K}$ (in Eq. (1.1)) as samples to estimate of the covariance matrix as: $\mathcal{K}_i = \frac{1}{K-1} \sum_{k=1}^K (x_i^k - \bar{m}_i)(x_i^k - \bar{m}_i)^T$, with $\bar{m}_i = \frac{1}{K} \sum_{k=1}^K x_i^k$. To estimate the determinant of a covariance matrix, the product of its non-zero singular values (SVD) is used in order to avoid numeric issues (i.e. the pseudo-determinant is used instead of the determinant).

V. ELECTRIC VEHICLE CHARGING SCENARIO

We exemplify the proposed framework in a simulated scenario similar to [6], but using a simpler generative model for the local cost. Our goal is to analyze the effect of two factors related to the hierarchical distributed coordination: *i*) what kind of grouping is obtained using the proposed flexibility measure, and *ii*) how well does the obtained hierarchical distributed structure can adjust to unplanned changes. In particular we analyze a difficult case: we assume there are communication problems between each group and the global coordinator, and thus only coordination within each group is possible.

Initial setup and EV modeling: The scenario consists of a community of $N=256$ households, each having an EV that must be charged (3kWh) continuously for 3 hrs during the day. A day is divided into $T = 144$, 10-minute slots, thus the charging takes 18 time-slots. Each EV is represented by three modes ($M = 3$): $\mathcal{Q} = \{q_1, q_2, q_3\}$, indicating the period before, during, and after charging. In this scenario, if the charging time is not coordinated, all EVs will be charged at a similar time (e.g. in the evening after the EV is plugged) and generate a large peak. Thus it would be better that the EVs coordinate their charging time⁴.

We fix the duration of the charging time $d_2 = 18$ (second mode), and assume that: *a*) agent i has a preferred charging starting time μ_i defining the distribution of the duration of mode q_1 , *b*) the control variable is the starting time u_i , and *c*) the starting probability is given by a Gaussian centered in μ_i and with variance σ_i^2 . Given that we assume that the duration of the charging is fixed, and by also assuming that the duration of the last mode (q_3) follows a uniform distribution, we have that $f_i^u(u_i)$ equals $\frac{(u_i - \mu_i)^2}{2\sigma_i^2} + c_i$, with σ_i^2 defining how flexible household i is, and c_i a non-relevant constant. For the different EVs we set a preferred starting time μ_i uniformly distributed in [50, 70], and let σ_i take values in $\{3, 9\}$ (evenly distributed),

⁴We consider an EV charging scenario for simplicity and for reducing the number of parameters (we focus here on the hierarchical structure, not on the modeling of appliances). However, any controllable appliance can be included, e.g. lighting (level adjustment), A/C and laundry machines (usage timing), etc.

i.e. about half the users are flexible in their start charging time, and the other half is less flexible.

We assume that each EV requires $1000[W]$ when being charged, which happens during mode q_2 . This is modeled by considering an output distribution given by $P(x_{i,t}|s_t) = 1_{[x_{i,t} \in \mathcal{X}(s_t)]}$, with $1_{[0]}$ the indicator function, $x_{i,t}$ the profile x_i at time t , and $\mathcal{X}(s_t) = 0, 1000, 0[W]$ for $s_t = q_1, q_2, q_3$ respectively. Thus $f_i^{x|u}(x_i, u_i) = -\sum_{t=1}^T \log 1_{[x_{i,t} \in \mathcal{X}(s_t)]}$, and x_i is uniquely determined by u_i .

Day-ahead coordination: We assume that the households together with a global coordinator perform a day-ahead coordination as in Fig. 1. In this day-ahead coordination, the community tries to flatten its aggregated power profile using a global cost $g^0(\sum_i x_i) = \alpha \|\sum_i x_i\|^2$ using the negotiation presented in Eq. (1). The aggregated profile obtained from this negotiation is the reference $r = \sum_i r_i$ for the day. In Fig. 3 (blue) the day-ahead coordination result is shown, together with the power usage profile that would be required if there is no coordination (red) between the households.

Grouping: After the day-ahead negotiation takes place, the global coordinator uses the observed history of the negotiation $\{x_i^k\}_{k=1}^K \forall i$ to estimate the covariance matrices of all agents \mathcal{K}_i , and then partitions the agents into $J = 4$ groups seeking to maximize the flexibility of the groups by applying the *uniform assignment algorithm* using 10^4 sampled allocations. In Fig. 5 the obtained covariance matrices of two EVs and the community are shown, while in Fig. 6 the covariance matrices of the four obtained groups are shown. Note that there are negative entries in the covariance matrices (off diagonal), due to the negative correlation between start and end charging times. The obtained aggregated reference profile for each groups, according to the day-ahead aggregated profile, is presented in Fig. 4.

On-line coordination: Now we assume that at time $t = 10$, the agents in set $\hat{\mathcal{N}} \subseteq \mathcal{N}$ change their preferred starting time, and the same agents also reduce their flexibility. We consider different ratios for the number of such agents, defined by $p = |\hat{\mathcal{N}}|/|\mathcal{N}|$ with $p \in \{0, 0.2, 0.4, 0.6, 0.8\}$. Given that some agents change their cost function (i.e. their power usage profile), a new coordination takes place considering the new cost function of the EVs $i \in \hat{\mathcal{N}}$, with the updated parameters (σ_i, μ_i) set as: $\sigma_i \leftarrow 1$ and $\mu_i \leftarrow \mu_i + \delta_i$, where δ_i is uniformly distributed in $[-10, 10]$, i.e. the vehicles change their preferred starting time, and reduce considerably their flexibility (σ_i changes from 3 or 9 to 1).

In the day-ahead coordination, the goal of the community was to flatten the power usage, however now the goal of the community becomes to minimize its deviation from the day-ahead reference aggregated power usage $r = \sum_i r_i$ by using the global cost $g_r(\sum_{i \in \mathcal{N}} x_i) = \|\sum_{i \in \mathcal{N}} x_i - r\|^2$. We also assume that in some cases a group can not communicate with the global coordinator. In those cases the group coordinator will add a group cost function (as in Eq. 5), $\hat{g}_j(\sum_{i \in \mathcal{N}_j} x_i) = \|\sum_{i \in \mathcal{N}_j} x_i - \sum_{i \in \mathcal{N}_j} r_i\|^2$, seeking to minimize the deviation from the group's day-ahead reference profile.

Figure 7 present results of the coordination when no group can communicate with the global coordinator (i.e. simulating communication problems), when a ratio p of the agents change

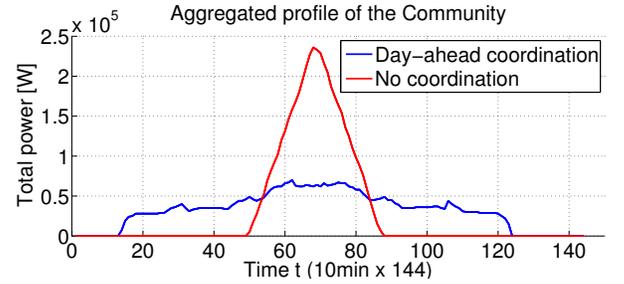


Fig. 3. Day-ahead coordinated profile of the community (blue) and non-coordinated profile of the community (red). Thanks to the coordination, the community flattens its profile and cuts the maximum to a third.

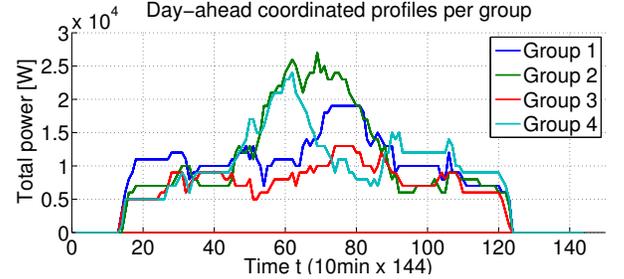


Fig. 4. Reference profile for each group according to the day-ahead coordination (their sum corresponds to the coordinated profile in Fig. 3).

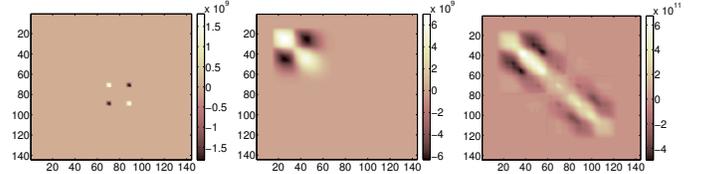


Fig. 5. Covariance matrices used to calculate the entropy. *Left to right:* EV with $\sigma = 3$, EV with $\sigma = 9$, and all EVs (community). Note the negative correlations between start and end charging times.

their profiles. Fig. 8 summarizes the results for different values of p and when different numbers of groups are able to communicate with the global coordinator.

Comments: The grouping results (Fig. 6), together with the day-ahead reference profiles for those groups (Fig. 4), show that the obtained groups have some flexibility for most time slots, that in some cases there is concentration around some time slots (e.g group 4), and that in other cases – e.g group 3 – the covariance matrix is smoother, has less extreme negative/positive values and the group's aggregated profile is more flat. The coordination results (Fig. 7 and Fig. 8) show that even when 20% or 40% of the agents change their planned pattern usage, the groups are able to adjust and obtain a profile similar to their reference, even if there is no coordination among groups (simulating communication issues). More over, even when 80% ($p=0.8$) of the agents change their profile and become inflexible, and there is only coordination within the groups (Fig. 8, red curve) the total aggregate profile is far from the no-coordination case.

VI. DISCUSSION AND CONCLUSIONS

We presented a distributed hierarchical coordination framework for balancing the power usage of a community. The

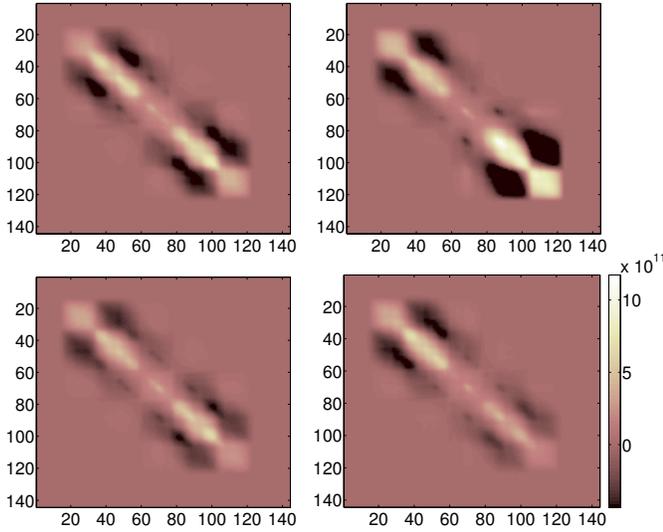


Fig. 6. Covariance matrices used to calculate the entropy of the groups. From top-to-bottom, left-to-right: groups 1, 2, 3 and 4. Same color map is used for all groups. Note the negative correlations between start and end charging times.

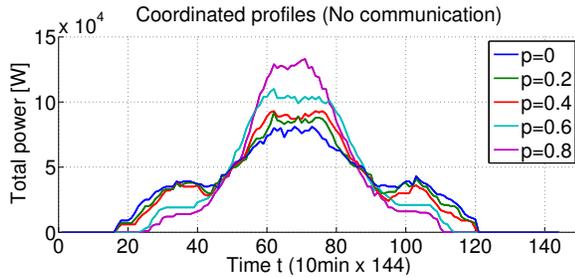


Fig. 7. Community aggregated profiles when the groups can not communicate with the global coordinator. The value of p indicates the ratio of agents that change their cost function and become inflexible. E.g. a value of p equal to 0.2 means that 20% of the agents change their cost function.

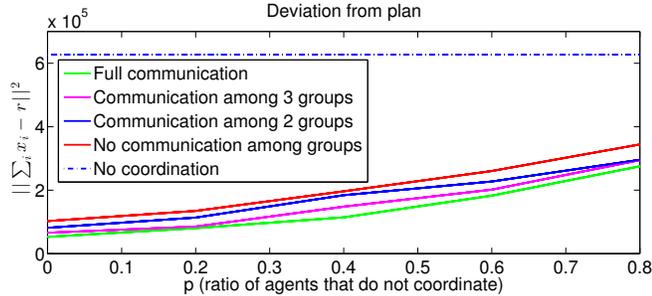


Fig. 8. Deviation from reference when a set of groups can not communicate with the global coordinator and $p \cdot 100\%$ of the agents do not follow the plan. The *no coordination* case (blue dotted line) is shown for comparison.

community can plan a day-ahead usage pattern, and during the day, coordinate to stay close to the planned pattern, allowing to reduce the generation cost, add predictability to their power usage, and help giving stability to the grid. The hierarchical architecture is built by allocating the agents in groups, seeking that each group is flexible enough to adjust to unplanned changes, even when there is no coordination with upper levels of the hierarchy. We propose the use of a submodular welfare maximization framework for grouping the agents, and define

the flexibility of a group in terms of the entropy of the group's aggregated profile. During the grouping, other welfare measures can be included in addition to the flexibility. Now we briefly comment two options: *i*) For reducing cost in the deployment of the communication network, it would be desirable that the agents in a group had a low geographical dispersion, to allow faster communication and higher robustness in case of failures. This would require defining a submodular dispersion measure, or including a non-submodular dispersion function as in [16]. *ii*) Two agents may overload a substation if they do not coordinate (e.g. as in [11]). Thus it would be better if they are in the same group (subtree of the coordination architecture). Therefore, information of the distribution network topology should be added to the welfare measure of a group.

Acknowledgment: This work was supported by CREST from the Japan Science and Technology.

REFERENCES

- [1] M. H. Albadi and E. El-Saadany, "Demand response in electricity markets: An overview," in *Power Engineering Society General Meeting, 2007. IEEE*, 2007, pp. 1–5.
- [2] S. Caron and G. Kesidis, "Incentive-Based energy consumption scheduling algorithms for the smart grid," in *First IEEE International Conf. on Smart Grid Communications (SmartGridComm)*, 2010, pp. 391–396.
- [3] M. Alizadeh, T.-H. Chang, and A. Scaglione, "Grid integration of distributed renewables through coordinated demand response," in *IEEE 51st Conf. on Decision and Control (CDC)*, 2012, pp. 3666–3671.
- [4] P. Palensky and D. Dietrich, "Demand side management: Demand response, intelligent energy systems, and smart loads," *IEEE Transactions on Industrial Informatics*, vol. 7, no. 3, pp. 381–388, 2011.
- [5] F. Rahimi and A. Ipakchi, "Demand response as a market resource under the smart grid paradigm," *IEEE Transactions on Smart Grid*, vol. 1, no. 1, pp. 82–88, June 2010.
- [6] H. Kawashima, T. Kato, and T. Matsuyama, "Distributed mode scheduling for coordinated power balancing," in *IEEE Int. Conf. on Smart Grid Communications (SmartGridComm)*, 2013, pp. 19–24.
- [7] A. Krause and D. Golovin, "Submodular function maximization," in *Tractability: Practical Approaches to Hard Problems*, L. Bordeaux, Y. Hamadi, and P. Kohli, Eds. Cambridge Univ. Press, March 2014.
- [8] J. Vondrak, "Optimal approximation for the submodular welfare problem in the value oracle model," in *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, ser. STOC '08. New York, NY, USA: ACM, 2008, pp. 67–74.
- [9] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, Jan. 2011.
- [10] M. Ostendorf, V. Digalakis, and O. A. Kimball, "From hmms to segment models: A unified view of stochastic modeling for speech recognition," *IEEE Tran. on Speech and Audio Processing*, vol. 4, pp. 360–378, 1995.
- [11] O. Ardakanian, C. Rosenberg, and S. Keshav, "Distributed control of electric vehicle charging," in *4th Int. Conf. on Future Energy Systems (e-Energy '13)*. New York, NY, USA: ACM, 2013, pp. 101–112.
- [12] E. Shuiabi, V. Thomson, and N. Bhuiyan, "Entropy as a measure of operational flexibility," *European Journal of Operational Research*, vol. 165, no. 3, pp. 696 – 707, 2005.
- [13] D. Yao, "Material and information flows in flexible manufacturing systems," *Material Flow*, pp. 143–149, 1985.
- [14] M. Madiman, "On the entropy of sums," in *Information Theory Workshop, 2008. ITW '08. IEEE*, May 2008, pp. 303–307.
- [15] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*, 99th ed. Wiley-Interscience, Aug. 1991.
- [16] A. Dasgupta, R. Kumar, and S. Ravi, "Summarization through submodularity and dispersion," in *ACL (1)*. The Association for Computer Linguistics, 2013, pp. 1014–1022.