

Correlation among smart grid actors

Nicolas Gensollen, Monique Becker, Vincent Gauthier and Michel Marot

CNRS UMR 5157 SAMOVAR,

Telecom SudParis/Institut Mines Telecom

Email: {nicolas.gensollen, vincent.gauthier, michel.marot, monique.becker}@telecom-sudparis.eu

Abstract—

I. INTRODUCTION

Designing stable power systems is a classical engineering challenge since blackout events can have catastrophic consequences. Nevertheless, stability is a general concept and can be studied from several points of view. A recent approach introduced by complex systems theorists consists in abstracting the power grid as a graph where nodes represent loads, generators, or transformers, while edges stand for electrical lines. As graphs, power grids display some characteristics like clustering coefficients, degree distributions, or density for instance, that enable scientists to get a better understanding of the structure of such systems. It appears that power grids are particular systems compared to well-known large graphs like the internet, the world wide web, or social networks. That is, they exhibit different characteristics, suggesting that they tend to behave differently. A common hypothesis for such a difference is that power grids are completely designed by a very small number of entities (electricity operators mainly), whereas other pre-cited graphs results from the interactions of billions of independent entities. Therefore, meaningful properties like power law degree distributions, preferential attachment, or hub oriented attacks do not apply with power grids.

Actually, considering only the underlying infrastructure when abstracting the power grid in a graph generally gives unrealistic results since they are mostly dynamical systems transporting electricity. As a second step, this dynamical component needs to be modeled using the physical laws governing the electricity behavior like the famous Kirchoff's laws. Besides, it is known that frequency synchronization in the grid is a necessary condition for stability. A nice solution abstracting this uses network models of coupled oscillators. For example, extended Kuramoto models using a swing equation based dynamic provide interesting results.

Understanding power grids seems like a reasonable starting point for studying so-called *smart grids* that gained a lot of attention recently. As there exists many good articles presenting smart grid ideas, concepts, or architectures, we will not dwell too much on this point. In this paper, we rather concentrate on a different approach to the stability concept. We consider a set of agents that are equipped with distributed energy resources (DER) as well as electrical loads. In other words, an agent can produce and consume energy at any time. His production can be of course used to met his own demand, but in cases where he is over-producing, we consider that he

has the possibility to sell his extra-production on a market. Such an agent model is known as a "*prosumer*" (and will be called accordingly in this paper).

Since the number of prosumers in a system may be large it seems unrealistic to have a communication topology where a central agent coordinates all the others so they can inject power whenever they are over-producing. Furthermore, as the production component of these prosumers is supposed to come mainly from renewables (wind and solar power for instance), this over-producing state might be rather unstable. Just imagine how strong the impact of a simple cloud passing in the sky on the agent production will be. It has been suggested that energy diversification and geographical expansion are stabilizing factors for the production. Moreover, forming coalitions of agents in order to achieve a goal, a popular approach in game theory, could be of good help for solving both problems. The system operator would only have to communicate with a relatively small number of aggregators/coalitions, that in turn will use some internal communication protocol to manage the involved agents. Such coalitional and hierarchical concepts are often referred to as virtual power plants (VPP) in the smart grid literature. The second point raised above is the central topic of the present paper : given N prosumers, what coalitions should be formed so that this over-production state has a high probability of being stable over the next period of interest ?

We will see that variability in the coalitions productions can be quantified to a certain extent by the correlation among the agents forming the coalitions. Understanding the correlation relationships among the agents can thus be illuminating in the sense that it will help us decide both what coalitions to form and how much they should sell. More precisely, we will build a framework in which the system operator has the possibility of fixing some entrance conditions on the market both in terms of stability and sufficient production. Based on predictions, coalitions can decide whether or not to enter and what power quantity they are willing to provide. Of course, coalitions failing at fulfilling their obligations during the contracted period of time (because of bad forecasts, unexpected rare events, or lying) will be exposed to financial penalties.

The paper is organized as follows, section 2 will give a brief overview of the related literature, section 3 will clarify how we generated realistic prosumer production traces based on weather data. In section 4, we will define most of the notations and explain why correlation between prosumer is a quantity of interest for our objective. Based on the conclusions of section 4, section 5 shows how we approached the problem

in a complex system fashion. Finally, section 6 will provide some results both on performance of the method and resilience of the coalitions formed.

II. GENERATING REALISTIC PROSUMER PATTERNS

An essential component of the smart grid is the smart meter which makes the interface between the end user and the rest of the system. Smart meters coupled with sensors measure quantities of interest like instantaneous consumption, receive informations from the grid (electricity price for instance), and take actions accordingly (demand side management program). Smart meters are currently and gradually deployed, and will probably provide interesting datasets to work on. Unfortunately, at the time this paper was written, production and consumption data for prosumers over a large region were not yet available to our knowledge. Some interesting experiments are notwithstanding being conducted and data are progressively made public (see for instance the ISSDA experiment in Ireland).

We denote by $P_i(t)$ the instantaneous extra-production of agent i at time t . It can be expanded in $P_i(t) = \bar{P}_i(t) - P_i^D(t)$, where $\bar{P}_i(t)$ represents the total production of agent i at time t and $P_i^D(t)$ its consumption at time t . In other words, $P_i(t)$ represents the instantaneous power that agent i is willing to sell at time t . As explained above, since large datasets containing this quantity over time are not yet available, we simulated these traces by considering separately \bar{P}_i and P_i^D .

For a prosumer i , it is possible to write both quantities as a sum over the distributed energy ressource (DER_i) and loads ($load_i$) of i : $\bar{P}_i(t) = \sum_{k \in DER_i} P_k(t)$ and $P_i^D(t) = \sum_{k \in load_i} P_k(t)$. For simplicity, in this paper we only consider windturbines (WT) and photovoltaic panels (PV) as possible DER for the agents ($DER_i = WT_i \cup PV_i$): $\bar{P}_i(t) = \sum_{k \in WT_i} P_k(t) + \sum_{k \in PV_i} P_k(t)$.

We denote by $\nu_i(t)$ and $\xi_i(t)$ the wind speed (in $m.s^{-1}$) and the solar irradiance (in $W.m^{-2}$) at agent i location and at time t , so that :

$$\bar{P}_i(t) = \sum_{k \in WT_i} \mathcal{F}_{WT}(\nu_i(t)) + \sum_{k \in PV_i} \mathcal{F}_{PV}(\xi_i(t)) \quad (1)$$

Where \mathcal{F}_{WT} (resp. \mathcal{F}_{PV}) is the power curve for the windturbines (resp. photovoltaic panels). We made here the implicit assumption that all windturbines (resp. photovoltaic panels) have the same power curve. More details about power curves and their approximations can be found in [].

Weather quantities like wind speed appear thus as alternative data for generating the P_i series. Fortunately, these kind of data are easier to find, and since the developement of small personal weather stations, their granularity keeps increasing. The idea for generating the P_i series is pictured in the first block of the process diagram. Basically, the space is discretized into zones around the chosen weather stations. A simplifying assumption is that weather quantities are constant over their zone. That is, if prosumer i and j are positioned on the same zone, their are exposed to the same weather.

Note that a prosumer i is defined by his zone Z_i as well as the sets DER_i and $load_i$. That is, a prosumer can be configured to represent anything from a single windturbine for instance ($DER_i = \{WT_0\}$ and $load_i = \emptyset$) to a pure load ($DER_i = \emptyset$ and $load_i = \{L_0\}$) through more complex combinations. In practice, we use random configurations for the agents.

In the rest of the paper, we use french weather data (see www.infoclimat.fr) starting in january 2006 and ending in december 2012, with a sampling frequency of three hours, and generate N timeseries of extra-production over this date range.

III. RELATED WORK

IV. NOTATIONS

As explained in section 2, we have a set $\mathcal{A} = \{a_1, a_2, \dots, a_N\}$ of N prosumers configured randomly, and for each agent, we simulated his extra-production $P_i(t)$, $\forall i \in \mathcal{A}$ from 2006 to 2012. Our objective is now to form groups of prosumers (the coalitions) so that the global power production resulting from the superposition of individual's extra-productions be sufficiently predictable. A parallel objectiv is to avoid very large coalitions as they would provide poor ameliorations for the communication load. Let $P_S(t) = \sum_{i \in S} P_i(t)$ be the extra-production of coalition S at time t .

Suppose now that coalition S has to suggest a production value P^{CRCT} to enter the market. This means that, during the time S is on the market, it will have to inject in the grid exactly P^{CRCT} at any time t and will be rewarded proportionally to this amount, with penalties if it deviates. Obviously, the actual extra-production will not be constant at this value and will oscillate due to various reasons. If S always produces more than P^{CRCT} , it will never have to pay penalties, but it is losing some gains since it could have announced a higher contract value. If the production oscillates around P^{CRCT} , by using batteries or demand side management techniques, S could be able to maintain its production to the contract value at any time. Nevertheless, if the oscillations are too important compared to the available storage capacity, S will probably break the contract and pay penalties. We can see that there is a return over risk tradeoff here, meaning that coalitions should find the right balance between announcing too low and losing some potential gains, and claiming too high and paying penalties.

Let us illustrate the rest of the notations and concepts with a simple example. We consider only two agents i and j such that the distribution of their extra-production can be approximated by normal distributions : $P_i \sim \mathcal{N}(\mu_i, \sigma_i)$ and $P_j \sim \mathcal{N}(\mu_j, \sigma_j)$. This is only for explanation purposes as it is of course rather unrealistic in real situations where the distributions are skewed. Using simple statistics, we can write the distribution of the coalition $\{i, j\}$ as $P_{\{i,j\}} \sim \mathcal{N}(\mu_{ij}, \sigma_{ij})$, where :

$$\begin{cases} \mu_{ij} = \mu_i + \mu_j \\ \sigma_{ij} = \sqrt{\sigma_i^2 + \sigma_j^2 + \rho_{ij}\sigma_i\sigma_j} \end{cases} \quad (2)$$

ρ_{ij} being the Pearson's correlation coefficient between P_i and P_j . If the coalition $\{i, j\}$ proposes a contract value P^{CRCT} , all instants where $\{i, j\}$ will produce less than P^{CRCT} is critical. Indeed, in this situation, $\{i, j\}$ will either have to discharge batteries to keep up with its contract, or pay penalties to the grid. The probability that $\{i, j\}$ is underproducing compared to the contract : $Pr[P_{i,j} \leq P^{CRCT}]$ is thus an important indicator of the coalition's credibility on the market. A well-known result for normal distributions is that the cumulative distribution function can be written as :

$$Pr[P_{ij} \leq P^{CRCT}] = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{P^{CRCT} - \mu_{ij}}{\sigma_{ij}\sqrt{2}} \right) \right] \quad (3)$$

The amount of risks a given coalition is willing to take depends on a lot of things, among with its capacity to compensate for under-producing (using batteries, backup generators...). Selecting the right contract value appears thus as an interesting problem on its own that we plan to investigate in future works. In order to keep the present paper in a reasonable length, we simplify the contract value selection problem a little bit by giving some responsibilities to a third party named the grid operator. The role of the grid operator is to constrain the market entry with an absolute low bound on the contract value and a upper limit relative to the coalition. In other words, the grid operator restrict the market to coalitions able to propose both sufficiently high and sufficiently credible contract values. More formally, let $\phi \in [0, 1]$ be a threshold fixed by the grid operator as a maximum value for the probability of under-producing. The highest contract value that a coalition can propose is thus P^{CRCT} such that $Pr[P_{ij} \leq P^{CRCT}] = \phi$. In the gaussian example, this translates by coalition $\{i, j\}$ announcing :

$$P^{CRCT} = \mu_{ij} + \sqrt{2}\sigma_{ij}\operatorname{erf}^{-1}(2\phi - 1) \quad (4)$$

This is the best contract value that the coalition S can afford giving the stability policy ϕ of the grid operator, and it is computable by any coalition. In order not to overload the market with unrealistically small coalitions, the grid operator also specify a lower bound P^{MIN} on the contract values. We thus characterized a valid coalition as one satisfying the two conditions :

$$\begin{cases} Pr[P_{ij} \leq P^{CRCT}] \leq \phi \\ P^{CRCT} \geq P^{MIN} \end{cases} \quad (5)$$

The gaussian assumption of this small example is convenient as it allows us to write all conditions in closed forms. Nevertheless, such assumption is rather unrealistic in practice. In the following, we thus release this assumption and work with observed distributions.

V. COALITION FORMATION

Now that we defined the notions of contract values and valid coalitions, we can face the problem of forming the "right" coalitions given a pool of agents. In this paper, we consider the "right" coalitions as being the ones that maximize some utility notion that, basically, indicates how much stable power

can be injected in the grid. This section aims at formalizing this utility notion, while explaining how this utility will be optimized in a greedy fashion over the correlation structure of the agents.

A. Utility function

We designed our model in a way that coalitions are remunerated proportionnally to their contract values $\mathcal{U}(S) \propto P_S^{CRCT}$. That is, if λ is the unitary price rate for electricity, a coalition S injecting P_S^{CRCT} in the grid during a period $[t_0, t_k]$ earns :

$$\mathcal{U}(S) = \int_{t_0}^{t_k} \lambda P_S^{CRCT} dt = P_S^{CRCT} \lambda \Delta_t \quad (6)$$

Since λ appears, in this simple model, just a multiplicative constant, we consider for simplicity $\lambda = 1$ in the following. The utility of S is now the amount of energy S plans to inject during its contract period. Since we focused on power quantities from the begining of the present paper, and because the power injected by S in the grid is supposed to be constant over time, we simplify further the utility to its contract value : $\mathcal{U}(S) = P_S^{CRCT}$. Although concise, this formulation suffers a major drawback. It is indeed not a concave function of the coalition length, meaning that coalitions can grow as large as the number of agents allows it, without any counterparty.

Such a model, that virtually allows infinitely large coalitions, is in practice not realistic. There are indeed costs (communication costs for instance) that increase with the coalitions sizes. We take this observation into account by rescaling the utility of a coalitions by its size in term of number of agents :

$$\mathcal{U}(S) = \begin{cases} \frac{1}{|S|^\alpha} \frac{P_S^{CRCT}}{P_S^{MAX}}, & \text{if } S \text{ is valid,} \\ 0, & \text{if } S \text{ is not valid} \end{cases} \quad (7)$$

Where α controls to what extent the size of the coalition impacts its utility, and P_S^{MAX} is a normalizing factor. P_S^{MAX} is the maximum production possible, which only occurs when all consumptions are null and all generators produce their maximum power.

Based on \mathcal{U} , the marginal contribution of an agent i can be expressed as $\delta_S(i) = \mathcal{U}(S + \{i\}) - \mathcal{U}(S)$. A coalition S has thus an interest in adding an additional agent i if this marginal contribution is positive :

$$\delta_S(i) \geq 0 \Leftrightarrow P_{S+\{i\}}^{CRCT} \geq P_S^{CRCT} \left(\frac{|S|+1}{|S|} \right)^\alpha \frac{P_{S+\{i\}}^{MAX}}{P_S^{MAX}} \quad (8)$$

Clearly, if $\alpha = 0$, whatever the size of the coalition, additional agents are added as long as they increase the contract value. Values of α greater than zero implies that additional agents must improve the contract value by a certain coefficient that decreases and converges to one as the size of the coalition is growing. For instance, if $\alpha = 1$ and $|S| = 1$, an additional agent must first double the contract value to be included, the next agent will then have to increase it by a factor of 1.5, and so on... Obviously, α impacts directly the

sizes of the coalitions formed. In the following, we consider the case where $\alpha = 1$.

As can be pointed out, the purpose of \mathcal{U} is not a study of coalitions stability against player defection, which game theory provides a lot of tools for. This is indeed a problem on its own. The redistribution of a coalition's utility in terms of individual payoffs will therefore not be considered directly in this paper. \mathcal{U} can be rather interpreted as a measure of how good a given coalition is according to our criteria since it favors small coalitions with a good production to risk ratio.

B. Representing the correlation structure

The previous section explained that coalitions with small variances in their production probability distributions are more likely to afford high contracts. Furthermore, the gaussian example highlighted that the correlation structure of the agents plays an important role in finding stable coalitions. Usually, this correlation structure is formalized with a covariance matrix or a correlation matrix that contains all the correlation coefficients between the agents : $M = (\rho_{ij})_{\forall i,j \in \mathcal{A}^2}$.

Since the work of Mantanega, an interesting approach consists in computing a distance metric based on the correlation coefficients in order to organize the timeseries in a correlation graph where the weight of an edge between two timeseries (nodes) is the metric value for these series. In the following, we use two opposite distance metrics :

$$\begin{cases} d_{ij}^1 = 1 - \rho_{ij}^2, \\ d_{ij}^2 = \rho_{ij}^2 = 1 - d_{ij}^1 \end{cases} \quad (9)$$

Clearly, d^1 (resp. d^2) maps two correlated series as close points (resp. distant) while two uncorrelated series are distant (resp. close). These metrics enable us to compute a correlation graph $G_1 = (\mathcal{A}, E_1)$ and a "de-correlation" graph $G_2 = (\mathcal{A}, E_2)$, where E_1 is considered different from E_2 since the weights are different.

Because the distance metrics can be computed for all pairs, these graphs are complete and of little use as is. Historically, the approach used by Mantanega was to compute a minimum spanning tree over the correlation graph as to extract a correlation structure of the form of a hierarchical clustering. Later on, it was pointed out that, by definition, a spanning tree could not capture the underlying clustering structure hidden in the correlation graph. Another filtering approach by mean of a threshold ϵ on the edge weights solves this problem and allow one to exhibit clusters of correlated series.

Selecting the right filter ϵ is an important point since it affects the correlation structure the coalition formation algorithm will be working on. Unfortunately, there seem to be no clear consensus in the litterature on how to select such a threshold. For our purpose, we have a number of coalitions N_{COAL} that we want to generate from cliques in the graph, we need therefore at least N_{COAL} cliques of a given size to start. Besides, we consider coalitions as disjoint sets in this section, meaning that the starting cliques must be non overlapping. We thus select the optimal threshold as :

$$\epsilon^* = \min_{\epsilon \in [0,1]} \{ \epsilon \text{ s.t. } |\Theta_k(G_2^\epsilon)| \geq N_{COAL} \} \quad (10)$$

Where G_2^ϵ is the de-correlation graph G_2 filtered by ϵ , and $\Theta_k(G)$ is the set of non overlapping cliques of size k in a given graph G . In other words we select ϵ as the smallest threshold possible such that the filtered de-correlation graph contains at least N_{COAL} non overlapping cliques of size k . If the set is empty, it means that either N_{COAL} or k is too large.

Correlation can be seen as cosine of angles in L^2 , hence even if there is no strict transitivity relation for correlation, there is, to a certain extent, some partial notion of it. More precisely, if a , b , and c are three items such that $\rho_{ab} > \delta$ and $\rho_{bc} > \delta$, then we know, by the cosine addition formula¹, that $\rho_{ac} > 2\delta^2 - 1$. That is, if a and b are strongly correlated ($\rho_{ab} > 0.9$) and b and c are also strongly correlated ($\rho_{bc} > 0.9$), then there is a high probability for a and c of being strongly correlated ($\rho_{ac} > 0.62$). This is one of the reasons why searching for clusters in correlation graphs, that is, clustering according to the correlation, makes sense.

Nevertheless de-correlation seems like a more complex concept than correlation in the sense that there is not even a partial notion of transitivity when it comes to it. As expected, the clustering coefficients of G_1 is much higher than the one of G_2 . This can be seen as another formulation of Onnela's study on the structural roles of weak and strong links on correlation graphs. Strong links, accounting for strong correlation relationships, are responsible for the clustering, while weak links provide the connectivity between clusters. Searching for clusters in G_2 and hoping that this strategy will provide a nice coalition structure of internally uncorrelated coalitions seems thus pointless.

Consider now a clique in G_2 , which is a complete subgraph of G_2 . This is indeed a structure of interest for our purpose. Since there is a link for every pairs of nodes, we know, by construction, that a clique has a mean correlation and a maximum correlation less than ϵ . More formally, let

$$\begin{cases} \bar{\rho}_S = \frac{2}{|S|(|S|-1)} \sum_{i \in S} \sum_{j \in S, j > i} \rho_{ij}, \\ \rho_S^> = \max_{(i,j) \in S^2} (\rho_{ij}) \end{cases} \quad (11)$$

be the mean correlation of coalition S and the maximum correlation value in S . If S is a clique in G_2 and $\epsilon \in [0, 1]$ is the threshold used for filtering G_2 , then $\bar{\rho}_S \leq \frac{2}{|S|(|S|-1)} \sum_{i \in S} \sum_{j \in S, j > i} \epsilon$, that is, $\bar{\rho}_S \leq \epsilon$, and $\rho_S^> \leq \epsilon$.

Because of this de-correlation property, cliques in G_2 appear as good candidates for coalitions. Nevertheless, by doing so, coalitions are often small and only formed based on the underlying correlation structure of the agents, and not on power generation capacity. It is possible that adding agents to these coalitions has the combined effect of increasing the production while decreasing its stability. The question revolves around measuring the benefits of this production surplus compared to the disadvantage of a higher instability. Hopefully, this can

¹ $\cos(a+b) = \cos(a)\cos(b) - \sin(a)\sin(b)$

be quantified exactly with the marginal benefit developed in the previous section.

C. Coalition formation algorithm

The algorithm takes inputs from :

- **The agents** : historical series of available productions P_i ,
- **The grid operator** : market entrance policy (P^{MIN}, ϕ) ,
- **The "user"** : Number of desired coalitions N_{COAL} and size of starting cliques k .

The first steps consists in computing the de-correlation graph G_2 as well as the optimal threshold ϵ^* . Cliques of size k in $G_2^{\epsilon^*}$ are considered as coalition seeds. The next step is a local greedy improvement over the correlation structure represented by $G_2^{\epsilon^*}$. Cliques add alternatively the node i^* in their neighborhood that yields the best marginal benefit $MAX_{i \in N(clique)} \delta_{clique}(i)$ where $N(clique)$ is the neighborhood of a given clique. This addition occurs only if i^* is not already involved in another coalition, and if $\delta_{clique}(i^*) \geq 0$, meaning that utilities are increasing. The algorithm stops when all nodes are distributed in a coalition or when the global utility stops increasing :

Data: P_i series,
Grid policy (P^{MIN}, ϕ) ,
Desired number of coalitions N_{COAL} ,
size of starting cliques k
Result: $CS = \{S_1, \dots, S_{N_{COAL}}\}$
Compute $G_2^{\epsilon^*}$;
Find the N_{COAL} cliques in $G_2^{\epsilon^*}$;
while $U(CS)$ is improving **do**
 for each clique do
 Find i^* ;
 if $\delta_{clique}(i^*) \geq 0$ **then**
 $clique = clique \cup \{i^*\}$;
 end
 end
end

Algorithm 1: Algorithm

VI. RESULTS

The algorithm presented in the previous section is supposed to generate a given number of coalitions that have good utilities, and therefore, high contract values for relatively small sizes. Nevertheless, as it comprises mainly of a greedy optimization based on local improvements, the probability that the algorithm finds the optimal coalitions set is very low. Actually, it is not obvious that a strict optimum exists at all. Besides, there is no, to our knowledge, state of the art algorithm that aggregate uncorrelated series in an optimum way.

In order to have an idea about the algorithm quality, we compare its results with :

- **Random** : This algorithm is a random search over the coalition structures space. It analyses a given number of structures and returns the best that it has encountered.

- **Correlated** : This is the complete opposite of our algorithm. It basically uses the correlation graph G_1 and performs community detection on it. The resulting coalitions have thus very high internal correlations. We thus expect this algorithm to perform very bad compared to the others.

- **K-means** : This is simply a K-means clustering of the agents P_i 's series. The goal of presenting this algorithm is to show that K-means do not consider correlations explicitly and is therefore not well suited for our purpose.

The coalitions formed according to these algorithms will be compared using two criteria : the normalized global utility of the coalitions formed $\bar{U}(CS) \in [0, 1]$, and the percentage of valid coalitions that satisfy the grid conditions $\tau(CS) \in [0, 1]$:

$$\bar{U}(CS) = \frac{1}{N_{COAL}} \sum_{S \in CS} U(S) \quad (12)$$

$$\tau(CS) = \frac{|\{S \in CS, s.t. U(S) > 0\}|}{N_{COAL}} \quad (13)$$

VII. CONCLUSION