Coalition Formation Algorithm of Prosumers in a Smart Grid Environment

Nicolas Gensollen, Vincent Gauthier, Michel Marot and Monique Becker
CNRS UMR 5157 SAMOVAR,
Telecom SudParis/Institut Mines Telecom
Email: {nicolas.gensollen, vincent.gauthier, michel.marot, monique.becker}@telecom-sudparis.eu

Abstract—In a smart grid environment, we study coalition formation of prosumers that aim at entering the energy market. It is paramount for the grid operation that the energy producers are able to sustain the grid demand in terms of stability and minimum production requirement. We design an algorithm that seeks to form coalitions that will meet both of these requirements: a minimum energy level for the coalitions and a steady production level which leads to finding uncorrelated sources of energy to form a coalition. We propose an algorithm that uses graph tools such as correlation graphs or clique percolation to form coalitions that meet such complex constraints. We validate the algorithm against a random procedure and show that, it not only performs better in term of social welfare for the power grid, but also that it is more robust against unforeseen production variations due to changing weather conditions for instance.

I. INTRODUCTION

One of the key ideas in the smart grid revolves around the introduction of communication means inside the power grids. This could enable complex improvements in the energy management and leads progressively to a greener energetic system [1], [2]. Distributed energy resources (DER) such as wind turbines or photovoltaic panels are not supposed to emerge only in remote farms, but also in residential areas. Together with electric vehicles, and demand side management programs [3], they will constitute the building blocks which will help to turn the today pure energy consumers into true actors of the grid operation [1]. Such agents that both consume and produce energy are ready to make concessions (appliances delays, V2G) to ensure grid stability, and are commonly called prosumers [1], [4].

There seems to be a clear consensus on the benefit of having bidirectional communication flows between the prosumers and the grid, as demand side management relies on such an architecture [2]. Nevertheless as the number of active prosumers is expected to rise, it is safe to assume there is a need for more complex communication patterns (prosumers grouping together into so-called “coalitions”) that will help to decrease the communication burden and satisfy the multiple requirements of the power grid management. Formation of coalitions inside a smart grid environment could be applied to various types of agents such as self sufficient micro-grids [5], sizable and adjustable virtual power plants (VPP) [1], [6], or a fleet of electric vehicles that back up the grid in emergency situations (V2G) [1]. These are only a few use cases where the coalition of multiple agents can enhance the grid reliability and in the mean time reduce the communication burden.

In this paper, we consider the case where prosumers wanted to be clustered together into coalitions in order to be able to produce enough power which may be sold to the grid. If the coalition’s production is sufficiently high and stable, the coalition can negotiate on a day by day basis with the grid by announcing each day an expected production for the next day. We focus on how to statistically improve the production stability by carefully forming coalitions of prosumers that have greater probabilities of staying in acceptable range of production. By using prediction techniques, we form coalitions that meet the contract values ranges proposed by the grid, enabling it to schedule its production accordingly. Meeting contract values for the prosumers is especially relevant in power trade conditions, where energy is traded based on day-ahead predictions. In this context, participants should try to minimize their prediction errors in order to maintain a stable state for the grid. They may even occur some penalties if the productions deviate from the initial contract values. However, it remains pertinent for both side to maintain the production as stable as possible : for the prosumer, the stability (with renewable energy sources) will impact its net gain, and for the power grid, it will help to maintain its reliability.

In this paper, we seek to form coalitions that are able to announce "high contract values with high reliability". We thus aim at:

- Building a realistic prosumer model with renewable production sources based on weather data (see section III-A).
- Define a coalition formation model that enables the grid to set some requirements under which any group of prosumers will be allowed to sell its production (see section III-B).
- Define a utility function that will satisfy the grid requirements and maximize the stability of the prosumer coalition productions (see section III-C).
- Define a coalition formation algorithm (see section IV).

In section III we consider that the agent’s production depends on meteorological conditions (wind turbines and photovoltaic panels (PV) as generators), various energy mix and preferences, and their appliances (loads). We use meteorological data (see section III-A) to account for seasonal and daily variations in the prosumer’s energetic profiles as well as realistic geographical correlations between different agents.
We use a modified clique percolation algorithm (see section IV) that will enable us to expand the cliques as needed in order to form proper coalitions that fulfill both the grid requirements and lower the production volatility. Finally, section V provides some results and a conclusion in section VI.

II. RELATED WORK

Traditionally, forming coalitions in a pool of agents can be done either in a centralized way where a single central unit is responsible for all the computations or in a distributed way where agents have only local knowledge and take actions accordingly. It is of common use to represent the situation and assess the stability of the solution by using game theoretic tools. Some papers [8][9] focus then on finding an optimal coalition structure giving a pool of autonomous self interested agents using distributed merge/split algorithms.

Tackling the stability issues of renewable DER, the TradeWind project [10] simulated the impact of wind power on electricity exchange and cross-border congestions by using a flow-based market model. The idea revolves around identifying key european interconnections (already existant or not) in order “to make optimal use of the European spacial de-correlation of wind power”. It was indeed shown that geographical aggregation provides smoothening effects and that the amount of prediction errors for wind power in a geographical region diminishes as the region size increases, especially for short forecast horizons.

On a narrower scale, [11] study the formation, in a game theoretic setting, of virtual power plants (VPP) composed of multiple self-interested DER. Two requirements for the formation of virtual power plants are considered: the reliability of supply and the minimization of the entities the grid has to deal with. From this, [11] builds a pricing mechanism that encourages VPP to report true estimates of their aggregated production and penalizes prediction errors. A redistribution scheme of the VPP to the DER is also constructed such that the payoff allocation lies in the core of the game, meaning that no DER has an incentive to leave the coalition.

In this paper we do not consider stability against player defection but focus on a statistics-oriented definition of coalition stability. We want to form coalitions based on utilities that depend on statistical properties of time-series values. The setting is thus similar with some financial studies on stock exchanges, where researchers tried to find relevant correlated clusters of stocks based on their daily prices variations. There exists well-known algorithms such as K-means or hierarchical clustering that are traditionally used for such purposes. Nevertheless, despite their popularity, they do not appear particularly well suited for meaningfully translating complex correlation relationships between time series while clustering them, especially when these series exhibit complicated underlying patterns. In [7] the author introduced an approach where stocks time-series of their daily log returns, are organized in a graph such that stocks exhibiting similar price fluctuation patterns are close to each other. This closeness notion is formalized with a similarity measure based on Pearson correlation coefficient ($d_{ij} = \frac{1}{2} \sqrt{2(1-\rho_{ij})}$ or $d_{ij} = 1 - \rho_{ij}^2$) that enables to weight the edges of the graph.

\(\epsilon\)-graphs, consists in filtering edges based on their weight, only keeping edges whose weights are less than \(\epsilon\). In [12], [13], the authors studied the topological properties (average clustering, connectivity, relative number of cliques) of the correlation graph against those of growing random graphs, depending on the threshold \(\epsilon\). However, there is no well defined method to select a right tradeoff \(\epsilon\) as a function of the network topology.

Presented in this way, the time-series clustering task seems very close to graph community detection. Communities in networks are indeed often seen as groups of nodes exhibiting high internal densities of links as well as a low density across communities, and several topology oriented techniques for finding communities are present in the literature ([14][15][16]). For our purpose where decorrelation is the closeness notion, such algorithms tend to have some difficulties because only a few inclusions of very high correlations can strongly affect the stability of a coalition. However, local cliques, where all nodes are linked, provide uncorrelated groups of nodes. There exists greedy heuristics based on cliques in the literature such as clique percolation [17] which realizes local optimizations of a fitness function and results in overlapping community structures. Detection of overlapping communities is actually a very active field of research, especially in social networks where a person might belong to several communities.

III. MODEL

A. Prosumer model

Our major concern is to characterize a prosumer behavior with realistic patterns of consumption, production as well as realistic geographical correlations between them. Only wind turbines and PV will be used as generators throughout this paper, and we denote by \(A = \{a_1,...,a_N\}\) the set of prosumers. To account for the production of a prosumer, we rely on French’s weather data from 2006 to 2012 sampled every 3 hours [18] (similar data are available for the United States covering 2010 [19]). As shown in the first part of figure 1, the first step consists in discretizing the studied zone around well chosen weather stations and gathering three kinds of traces (wind speed, cloudiness, and temperature) that we organize in a structure we will refer to as "climate vector" in the following. We consider that climate vectors are constant over their area, meaning that, if two agents are in the same area, they are exposed to the same climate vector. We use existing DER models enabling conversion between weather variables and output powers (see power curves in [11][20]). [21] provides also a convenient way of using cloudiness traces as realistic degradation factors in a “clear blue sky” solar model.
cause they depend both on the energetic mix and the habits. Nevertheless, these macro variations can hide the interesting information in the correlation coefficients and twist the rest of our process. Thereby, we remove these seasonal effects for all agents (see second block of figure 1). We denote by \( T_i = \{ P_i(t_0), ..., P_i(t_K) \} \) the resulting time-series for agent \( i \), and by \( F(P_i) \) the corresponding probability distribution.

### B. Coalition formation model

Let us recall that our goal is to group prosumers into coalitions so that the global power production resulting of the superposition of individual prosumer’s productions be sufficiently predictable and above a minimum power level. We thus extend the agent notations to any coalition \( s \subset A \):

- \( P_s(t) = \sum_{i \in s} P_i(t) \)
- \( T_s = \{ P_s(t_0), ..., P_s(t_K) \} \) with \( F(P_s) \) the corresponding probability distribution

Furthermore, we consider that a coalition \( S \) has the possibility to announce a contract value \( P^{CRCT}_s \) on the market. Due to the volatile nature of DER productions or user consumptions, the available power of a coalition will oscillate around \( P^{CRCT}_s \). It is of the responsibility of the information system that forms coalitions to select agents in order to provide a production as stable as possible. The probability of \( s \) value being below its contract value \( Pr[P_s \leq P^{CRCT}_s] \) appears therefore as a quantity of interest that we would want to keep as low as possible.

To illustrate what is done in the following, let us consider a very simple example with two agents, say \( i \) and \( j \), with a production following a gaussian value probability distributions \( F(P_i) = \mathcal{N}(\mu_i, \sigma_i) \) and \( F(P_j) = \mathcal{N}(\mu_j, \sigma_j) \), such that the joint probability distribution \( \mathcal{F}(P_{ij}) \) of the coalition \( \{ij\} \) is also a gaussian with the following parameters:

\[
\begin{align*}
    \mu_{ij} &= \mu_i + \mu_j \\
    \sigma_{ij} &= \sqrt{\sigma_i^2 + \sigma_j^2 + \rho_{ij}\sigma_i\sigma_j}
\end{align*}
\]

\( \rho_{ij} \) the Pearson correlation coefficient between \( P_i \) and \( P_j \).

We can easily write the probability \( Pr[P_{ij} \leq P^{CRCT}_i] \) as:

\[
\frac{1}{2} \left[ 1 + \text{erf} \left( \frac{P^{CRCT}_i - \mu_{ij}}{\sigma_{ij}\sqrt{2}} \right) \right]
\]

Let \( \phi \in [0,1] \) be the reliability constraint imposed to the coalition \( \{ij\} \). It stipulates that if \( \{ij\} \) wants to join the market, the probability of \( \{ij\} \) value (at any instant \( t \)) being below its contract value should at most be \( \phi \). That is, \( Pr[P_{ij} \leq P^{CRCT}_{ij}] \leq \phi \). We also impose to \( \{ij\} \) a minimum contract value constraint denoted by \( P^{MIN}_M \), which states that the coalition should announce more than \( P^{MIN}_M \) to be accepted (\( P^{CRCT}_s \geq P^{MIN}_M \)).

Coalition \( \{ij\} \) has thus potentially a set of possible contract values that respects both rules. The strategy of \( \{ij\} \) (and all other coalitions in the paper) consists then in choosing, if it exists, the maximum value of this constrained set. We denote by \( P^{\phi}_s(s) \) this maximum value for a coalition \( s \). For
\(P(t) \in \mathbb{R} \quad (P_i(t))\)

- \(F(P_i)\) is the probability distribution of \(i\)’s production
- \(\phi \in [0, 1]\)
- \(p_{MIN} \in \mathbb{R}^+\)
- \(P^{C\text{R}CT}_s \in \mathbb{R}^+\)
- \(\mathcal{P}(s) \in \mathbb{R}^+\)
- \(\mathcal{U}_{\phi, p_{MIN}}(s)\)
- \(\sigma_{ij}\) is a utility of zero. Furthermore, it seems obvious that the utility

\[ \mathcal{U}_{\phi, p_{MIN}}(s) = \begin{cases} P_{\phi}(s) & \text{if } s \text{ valid} \\ 0 & \text{else} \end{cases} \]

Table I: Notations

| \(\{ij\}\) | \(P_{\phi}(s)\) can be computed using eq. 2: \(P_{\phi}(ij) = \mu_{ij} - \sigma_{ij}\sqrt{2\pi}f^{-1}(1 - 2\phi).\) If \(P_{\phi}(ij) \geq p_{MIN}\), then \(\{ij\}\) can announce \(P^{C\text{R}CT}_{ij} = P_{\phi}(ij)\) on the market, else it is not allowed to enter. It thus appears (as it was intuitively understandable) that, for equivalent sizes, coalitions with low relative standard deviations (\(\sigma_{ij}/\mu_{ij}\)) are able to announce higher contract values.

As explained in the example above, if a coalition \(s\) wishes to join the market, it has to choose a contract value within a constrained set. Only coalitions whose set is not empty are able to enter the market (valid coalitions). For simplification, we consider that valid coalitions will always apply the same economically consistent strategy of announcing the highest possible contract value of this set (\(P^{C\text{R}CT}_s = P_{\phi}(s)\)).

C. Utility function

Adressing the problem of eq 3 we choose a utility function (eq. 5) that derives directly from the above remarks. If a coalition cannot provide a valid contract value, it receives naturally a utility of zero. Furthermore, it seems obvious that the utility should increase with the contract value. The \(1/|s|\) term in eq. 5 indicates that we favor small coalitions, mainly because they are easier to maintain in terms of communications.

\[ \mathcal{U}_{\phi, p_{MIN}}(s) = \frac{P_{\phi}(s)}{|s|} \]
of non overlapping cliques in $G_\varepsilon$, we choose the $\varepsilon$ parameter as the smallest value such that $G_\varepsilon$ contains at least $N_{\text{COAL}}$ non overlapping cliques:

$$\varepsilon^* = \min_{\varepsilon \in [0,1]} \left\{ \varepsilon \mid s.t \ [\Theta(G_\varepsilon)] \geq N_{\text{COAL}} \right\}$$

We construct $G_\varepsilon$ such that it contains at least $N_{\text{COAL}}$ potential seeds that may or may not be above the grid requirements, but exhibit low correlations among the members. Naturally, grid requirements could eventually be reached and social welfare increased by incorporating more nodes into the seeds. For this purpose, we use a clique percolation algorithm on $G_\varepsilon$, that will extend the seeds over a correlation constrained environment. The main idea behind this algorithm is to make the seeds grow by local optimization of a fitness function, usually based on internal and external degrees of the seeds\(^\text{[17]}\). The main difference between our case study and this algorithm for community detection is that we chose the fitness function as our utility function (cf. eq\((5)\)) that both maximizes the production and minimizes its volatility. As explained in section\(\text{II}\) clique percolation leads generally to overlapping communities. For simplicity, in this paper, we wish to keep the coalitions disjoint and leave the management of overlapping coalitions for future work. We thus implemented a simple heuristic that considers nodes in multiple seeds one by one and chooses its final coalition as the one that “needs it the most” in terms of utility loss. More formally, for a coalition $s$ and a node $i \in s$, we define:

$$\tau_i(s) = \frac{U_{\phi, \ P_{\text{MIN}}}(s) - U_{\phi, \ P_{\text{MIN}}}(s - \{i\})}{U_{\phi, \ P_{\text{MIN}}}(s)}$$

If node $i$ belongs to multiple coalitions, the only coalition including node $i$ is the one that maximizes $\tau_i$. At this point, we have three degrees of freedom: the reliability ($\phi$), the required power to enter the market ($P_{\text{MIN}}$), and the number of desired coalitions ($N_{\text{COAL}}$). Next section shows the utility behavior within the parameters space $\{\phi, P_{\text{MIN}}, N_{\text{COAL}}\}$ and provides results of our 200 agents test.

V. RESULTS

As shown in figure\(\text{II}h\), the parameters $\phi$ and $P_{\text{MIN}}$ shape the utility function such that, if $\phi$ is close to zero, the reliability requirement is very high. Only small values of $P_{\text{MIN}}$ could then potentially lead to valid coalitions (and positive utilities). Conversely, the higher $\phi$, the less constraints are imposed to the coalitions and more valid coalitions can arise for a larger spectrum of $P_{\text{MIN}}$. Obviously, the highest utility values are found for high $\phi$, because the formed coalitions are able to announce higher contract values, yielding higher utilities.

In figure\(\text{II}h\), we fix the reliability to a given empirical value ($\phi = 0.08$). We can observe how the social welfare $\sum_{s \in S} U_{\phi, \ P_{\text{MIN}}}(s)$ evolves according to $P_{\text{MIN}}$ and $N_{\text{COAL}}$. In our setting we notice a maximum point around $N_{\text{COAL}} = 3$, followed by a decrease. There is clearly a maximum number of coalitions sustainable for a given set of parameters. After that value the coalitions newly formed won’t be able to pass the grid requirements. Moreover, reckon that increasing $N_{\text{COAL}}$ means also increasing $\varepsilon$, leading to denser graphs, meaning that the algorithm performance will also decrease.

In fig. 3 we perform a simulation with 200 prosumers, and we fix the reliability $\phi = 0.1$ meaning that coalitions should produce more than their contract values at least 90% of the time. We observe the result of our algorithm for a fixed number of coalitions of $N_{\text{COAL}} = 10$ and let the algorithm form the coalitions as the grid changes according to $P_{\text{MIN}}$. We compare our algorithm with a random process that splits the agents in $N_{\text{COAL}}$ coalitions. Fig. 3 shows this evolution for our algorithm (blue dots) and for the random process (red dots). The diameter of a dot is proportional to the number of agents in the coalition and the higher the dot, the higher its utility. The $P_{\text{MIN}}$ values of the $x$ axis are expressed in tenth of MW for readability and the hatched zone corresponds to the "under-requirement space", meaning that whenever a coalition is in this zone, it has a null utility. Looking at fig. 3, we can see first that our algorithm performs better than the random formation algorithm at finding high valued coalitions. The blue dots allowed to enter the market (non hatched zone) are indeed outnumbering the red dots, especially when the grid requirements are neither too low nor too high ($P_{\text{MIN}} = 5$ in fig. 3).
In more details, figure 4b presents the evolution of social welfare as the number of coalitions increases (all other parameters are kept constant) for random (red curve), clique percolation (blue curve), and correlated coalitions (green curves) that stands for a worst case scenario. As for figure 4b, it shows the percentage of coalitions able to enter the market for different values of $P_{MIN}$. For consistency, we average the results of both plots over 100 realizations (colored zones around the curves shows the standard deviations). When the grid requirements are constant (figure 4b), and the number of desired coalitions is low, clique percolation generally performs better than random case. Moreover, when $N_{COAL}$ gets bigger, the performance of a random split tumble down rapidly while our clique percolation is able to maintain efficiently the social welfare of the coalitions formed. When the grid requirements vary, for very low $P_{MIN}$, all coalitions for all algorithms are able to enter the market, yielding an acceptance percentage of 100%. But as $P_{MIN}$ increases, we see the percentage of the correlated coalitions is going down very quickly. After a few increases in $P_{MIN}$, the percentage of the random procedure starts dropping while it stays almost constant for our algorithm. For $P_{MIN} = 8$, we see that only a little more than half of the coalitions for the random case are able to enter while approximately 85% of them enters for the clique percolation algorithm. Finally, when the grid requirements becomes too high, the acceptance percentage of our algorithm tends slowly to zero (not shown in this plot for readability).

VI. CONCLUSION

We believe that the originality of this paper lies in its willingness to exploit the de-correlation of prosumer profiles in order to build stable coalitions. In this direction, we presented a model based on meteorological traces, that captures the complex "energetic mix/climate vectors" combination and generates realistic production and consumption patterns. We then built a framework that enables the grid to specify stability and minimum production requirements for filtering the coalitions. On this basis, we proposed a simple algorithm that seeks for uncorrelated prosumer patterns as potential seeds and expand them in coalitions able to rise above the grid requirements. We validated our algorithm against a random choice of coalitions, which can be considered as an average case. We showed that it performs better (coalitions are more stable and the overall production is more important) and that it exhibits a higher robustness/flexibility against grid requirements changes. The worst case scenario is represented by coalitions of correlated prosumers. Interesting leads for future works would be the use of correlated clusters to reduce the number of entities the algorithm has to deal with, or the introduction of a payoff allocation towards the prosumers such that the stability against player defection could be analyzed through game theory. Showing that maintaining the coalitions formed with our algorithm necessitates less communication and less storage capacity could also conduct to a stimulating project. Besides, we believe that not restricting the algorithm to non overlapping coalitions and studying the strategies and weights of nodes with multiple options could lead to interesting works.

REFERENCES

[1] S. D. Ramchurn et al., “Putting the ‘smarts’ into the smart grid: A grand challenge for artificial intelligence,” Commun. ACM, vol. 55, no. 4, pp. 86–97, Apr. 2012.
[2] C. Wu et al., Smart Grid Communications and Networking, E. Hossain et al., Eds. Cambridge University Press, 2012.
[3] P. Samadi et al., “Utilizing renewable energy resources by adopting DSM techniques and storage facilities,” International Conference on Communications, pp. 4221–4226, 2014.
[4] L. Xiao et al., “Anti-cheating prosumer energy exchange based on indirect reciprocity,” in International Conference on Communications, 2014, pp. 599–604.
[5] S. Palwa et al., “Topological Analysis and Mitigation Strategies for Cascading Failures in Power Grid Networks,” 2010.
[6] M. Braun, “Virtual Power Plants in Real Applications Pilot Demonstrations in Spain and England as part of the European project FENIX,” 2009.
[7] R. Mantegna, “Hierarchical structure in financial markets,” The European Physical Journal B - Condensed Matter and Complex Systems, vol. 38, no. 2, pp. 353–362, 2004.
[8] M. E. J. Newman, “Spectral methods for network community detection and graph partitioning,” p. 11, Jul. 2013.
[9] G. Dans et al., “Rayonnement solaire,” vol. 360, 2007.
[10] M. Girvan et al., “Community structure in social and biological networks,” Proceedings of the National Academy of Sciences, vol. 99, no. 12, pp. 7821–7826, 2002.
[11] M. E. J. Newman, “Community detection and graph partitioning,” no. 2, p. 5, May 2013.
[12] J.-P. Onnela et al., “Clustering and information in correlation based financial networks,” The European Physical Journal B - Condensed Matter and Complex Systems, vol. 38, no. 2, pp. 353–362, 2004.
[13] M. E. J. Newman, “Spectral methods for network community detection and graph partitioning,” p. 11, Jul. 2013.
[14] M. Girvan et al., “Community structure in social and biological networks,” Proceedings of the National Academy of Sciences, vol. 99, no. 12, pp. 7821–7826, 2002.
[15] M. E. J. Newman, “Community detection and graph partitioning,” no. 2, p. 5, May 2013.
[16] J.-P. Onnela et al., “Clustering and information in correlation based financial networks,” The European Physical Journal B - Condensed Matter and Complex Systems, vol. 38, no. 2, pp. 353–362, 2004.
[17] M. E. J. Newman, “Spectral methods for network community detection and graph partitioning,” p. 11, Jul. 2013.
[18] M. Girvan et al., “Community structure in social and biological networks,” Proceedings of the National Academy of Sciences, vol. 99, no. 12, pp. 7821–7826, 2002.
[19] M. E. J. Newman, “Community detection and graph partitioning,” no. 2, p. 5, May 2013.
[20] G. Dans et al., “Rayonnement solaire,” vol. 360, 2007.