Allowing Stable Groups of Prosumers to Enter the Market through Proper Aggregation

Nicolas Gensollen
Institut Mines Telecom
Computer Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332–0250
Email: http://www.michaelshell.org/contact.html

Vincent Gauthier
Twentieth Century Fox
Springfield, USA
Email: homer@thesimpsons.com

Michel Marot
and Monique Becker
Starfleet Academy
San Francisco, California 96678-2391
Telephone: (800) 555–1212
Fax: (888) 555–1212

***Abstract*—There is a growing interest for coalitional concepts in the smart grid, spanning multiple objectives such as power losses reduction, benefits maximization or market stability... In this paper, we study coalitions of prosumers (agents that produce but also consume depending on time) which aim at selling energy to the grid. We reckon that the grid is able to specify two requirements (stability and minimum production) that enable a coalition to enter the market. Based on probability distributions of units available power, coalitions which are able to achieve low standard deviations are considered as stable coalitions. As it seems intuitively understandable that maintaining stable coalitions would be less expensive in terms of communication load than highly versatible ones, our objective consists in forming coalitions that satisfy grid requirements both in terms of stability and minimum production. Our definition of stability being implicitely linked to correlation between agents timeseries (see section 2), we organize them over a correlation graph and use a variation of clique percolation to construct valid coalitions in reasonable time.**

I. Introduction

One of the most basic ideas in the smart grid revolves around the introduction of communication means in the power grids, which could enable complex improvements in the energy management and lead progressively to a greener energetic system [1] [2]. Distributed energy resources (DER) such as wind turbines or photovoltaic panels are supposed to emerge and populate not only spread out farms, but also classical residential neighborhoods. Together with electric vehicles, and demand side management tools, they constitute the building blocks which will help turn the today pure energy consumers into true actors of the grid operation [3]. Such energy aware agents that consume, produce, and are ready to make concessions (appliances delays, V2G...) to ensure grid stability are commonly called prosumers [4] [5].

It is rather clear that there should be communication flows between prosumers and the grid as most of the demand side management concepts suppose such a link [6] [7]. Nevertheless as the expected number of active prosumers in the future seems quite large, the literature assume more and more communication between prosumers as this is a key ingredient to maintain more complex structures (such as coalitions), and thereby decrease the communication load on the grid's side [8]. Furthermore, it has been shown that coalitions in the smart grid provide nice ways of improving stability while posing exciting challenges in terms of network infrastructure and management. Self sufficient microgrids that can disconnect from the main grid [9], virtual power plants (VPP) that map small DER in sizeable and adjustable plants [10], or coalitions of electric vehicles that back up the grid in emergency situations (V2G) [11] are just a few examples of how coalitional concepts can enhance grid reliability.

In this paper, our focus is on how coalitions of prosumers, aiming at selling energy to the grid, should be formed in order to support it efficiently. Namely, improving statistically production stability while decreasing communication needs for staying in stable state. The idea is that, by using prediction techniques, coalitions can propose contract values to the grid, enabling it to schedule production on its side accordingly. Actually, contract values are particularly relevant in power exchanges contexts where energy is traded based on predictions (day-ahead estimations for instance). It seems obvious that, in such systems, participants should try to minimize prediction errors in order to maintain a stable state of grid operation, and that some penalty rules should be applied to ensure that coalitions are willing to report contract values correctly. Furthermore, as the coalition productions are supposed to come from (uncertain) renewables, the notion of reliability comes into play with a visible tradeoff between contract values worth and reliability.

In this paper, our concern is on how prosumers should be aggregated in order to optimize this tradeoff, that is, how coalitions should be formed with the intention of announcing "high contract values with high reliability". We thus aim at:

- Realistic representations (prosumers productions are non independent variables and come (indirectly) from real traces)
- Minimizing the standard deviations of coalitions production probability distributions through the use of a proper utility function (see section 3)
- Being (relatively) scalable with the number of prosumers

In more details, and as will be explained in section 3, we consider agents that, depending on meteorological conditions, personal preferences, and their appliances (loads and generators), consume and produce more or less energy (using only wind turbines and PV as generators, which can be quite simply and efficiently modeled). We used real meteorological traces [13] to account for seasonal and daily variations in the prosumers energetic output profiles as well as realistic correlations. With these ingredients, we are able to run simulations and record the different output profiles as timeseries.

As correlations play a significant role in our mechanism, we organize the prosumers in a correlation graph, a quite popular approach in market stocks clustering [16] [17] (see section 2). We will see that in such graphs, cliques represent structures of interest. In our settings, it means that cliques are "good" (from the utility point of view) foundations for high valued coalitions. We thereby use a slightly modified clique percolation algorithm (see section 4) that will enable us to expand the cliques as much as needed in order to form proper coalitions that fulfill grid requirements.

The rest of the paper is organized as follows, section 2 locates our work in the related litterature. Section 3 explains the model we used to obtain realistic prosumers available power timeseries as well as some notations. Section 4 relates the method for building coalitions based on the agent timeseries. Finally, section 5 provides some results.

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 knowledges and take actions accordingly. There is a very nice and active smart grid litterature that covers the formation of coalitions trying to optimize various quantities (energy losses, communication bandwidth, cascade of failure sizes...). A common point being that a given quantity, usually transcribed by a utility function, should be optimized. In this paper, we restrict agents such that they do not take any actions in the coalition formation process. The focus is indeed not on the mechanism that drives the agents to an optimum structure, but rather on finding stable groups able to enter the power market for different levels of requirement. We will see in section 3 and 4 that, during the formation of coalitions, the agents are completely interchangeable with the timeseries representing their available production.

The setting is thus similar with some fiancial studies on stock exchanges, where researchers tried to find relevant clusters of stocks based on their daily prices variations. The problem of clustering is usually approached by means of a similarity measure and completed by a clustering technique such as K-means or hierarchical clustering. Nevertheless, in [] [], stocks are represented by the timeseries of their daily log returns. Quite intuitive similarity measures based on Pearson correlation coefficient are established ($d_{ij} = \frac{1}{2}\sqrt{2(1-\rho_{ij})}$ or $d_{ij} = 1-\rho_{ij}^2$) such that correleted stocks are close to each other. Such a situation can be represented by a complete graph $G = (V, E, \omega)$ such that a vertex stands for a stock and the weight of the edge connecting two stocks i and j is equal to d_{ij}. However, in this complete state, the graph G is of little use as precious information is flooded. Several filtering techniques have been proposed, among which, minimum spanning tree [], k-nearest graphs [], or ϵ-graphs [][].

The idea of ϵ-graphs consists in filtering edges based on their weight, only keeping edges such that their weight is less than ϵ. Despite the procedure simplicity, the choice of ϵ (or k for k-nearest graphs) is not trivial and may influence strongly the results. There is indeed to our knowledge no optimal rule for chosing ϵ. In [], the authors studied the topological properties (mainly average clustering and connectivity) of the correlation graph depending on the threshold ϵ. One of the conclusions was that "strong links" (i.e links between strongly correlated timeseries) are responsible for clustering while "weak links" provide network connectivity (in such a network, weak links are indeed connecting densely connected communities).

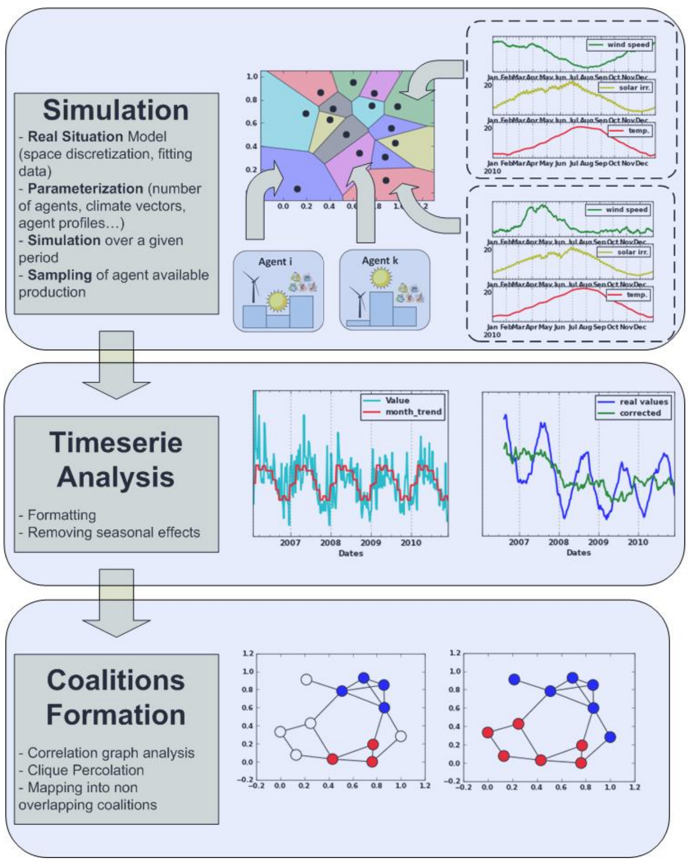
Presented in this way, the timeseries clustering task is also very close to graph community detection. Communities in networks are indeed often seen as groups of nodes exhibiting high internal densities of links as well as a low density across communities [1]. Although several techniques exists (based on different graph properties : spectrum of Laplacian, modularity, edge betweeness...), the clique percolation algorithm uses directly this observation and the idea behind it is actually quite simple [2] [3]. It starts indeed by searching for cliques (a complete subgraph) in the given graph and considers them as potential seeds for the different communities.

The next ingredient the algorithm needs in order to make the seeds grow is a fitness function that gives some value to a group of nodes. In community detection, it is of common use to employ topological functions that compare internal and external degree, but other functions can also be well suited depending on the problem. Seeds will then alternatively look in their respective neighborhood, select the node that yields the best increase in fitness, and finally incorporate it within the seed. The algorithm stops when seeds cease to grow or when all nodes are affected to at least one seed. At this point, some coalitions may be partially overlapping, which can be of great use especially in social networks where persons often belongs to several communities (family, friends, work colleagues...).

III. Model

A major concern while designing our model was its ability to translate correctly realistic patterns (consumption and production) as well as realistic geographical correlations between patterns. Largely because of these reasons (but also because they are easily, and sometimes freely, available online) we chose to use meteorological traces as inputs. In this paper we used namely both french data from 2006 to 2012 sampled every 3 hours (see link), as well as Western USA data over 2010 sampled every hour (see link). The weather stations already offer a quite good sampling of a given territory, and, with the growth of personal weather stations constantly updating data bases, the mesh becomes finer and finer. 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 traces for these stations. In this paper, we consider that prosumers can only produce through wind turbines and photovoltaic panels (PV). On the other hand, we built their consumption patterns according to two major cycles :

- Daily cycles : Consumption is low during night, and higher during the day with two picks in the morning and evening. Some noise is added so that prosumers have similar but not identical cycles.
- Seasonal cycles : Consumption is higher in the winter because of heating and low in the summer (we did not consider air conditioning in this paper). Temperature traces were the principal ingredient for modeling these cycles.



We thus collect for all chosen weather stations three kind of traces (average wind speed, cloudiness, and average temperature), which we will call a climate vector in the following. We consider that cimate vectors are constant over their area, meaning that, if two agents are on the same area, they are exposed to the same climate vector.

At this point, modeling agents consists in fixing a few parameters (most of the time drawn from random distributions) such as the geographical position, the number of wind turbines, PV, appliances, the temperature of confort, and so on... The objective is that, depending on the weather of her zone, the DER and appliances she owns, and the way she decides to heat her home, a given prosumer is able to compute her production/consumption at any time. We thus used simple and known power curve models for linking weather variables to output power, such that a wind turbine (or a PV for instance) takes a wind speed (a solar irradiance) as input and gives an instantaneous output power [] []. Any agent is thus able to compute at any time (within the traces bounds) how much she produces and consumes.

More formally, we denote by $\mathcal{A} = \{a_1, ...a_N\}$ the set of agents (the prosumers) and $P_i(t)$ the instantaneous power value of agent i at instant t (its instantaneous production minus consumption, meaning that $P_i(t)$ represents agent i available power at instant t). During the simulation (from t_0 to t_K), all agents record their values with an hour time interval. As expected (because we introduced them), the timeseries exhibit high seasonal patterns, but completely different from agent to agent because they depend both on the energetic mix and habits of a given prosumer. 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 $\mathcal{T}_i = \{P_i(t_0), ..., P_i(t_K)\}$ the resulting timeserie for agent i. For any agent i, we note $\mathcal{P}_i$ the probability distribution drawn from $\mathcal{T}_i$ and by P_i a random variable following $\mathcal{P}_i$.

We now extand these notations to any coalition $S \subset \mathcal{A}$:

- $P_S(t) = \sum_{i \in S} P_i(t)$
- $\mathcal{T}_S = \{P_S(t_0), ..., P_S(t_K)\}$ with $\mathcal{P}_S$ its probability distribution
- A coalition S has the possibility to announce a contract value P_S^{CRCT} on the market.

In this paper, we consider that coalition announcements are constrained by the grid. Namely, the grid has the possibility of fixing two parameters :

- the reliability (denoted by $\phi \in [0, 1]$)
- the minimum contract value (P_ϕ^{MIN}).

The reliability stipulates that for any coalition S willing to join the market, the probability of S value (at any instant t) being below its contract value should at most be ϕ. That is, $\forall S,\ \forall t,\ Pr[P_S(t) \leq P_S^{CRCT}] \leq \phi$. Furthermore, for consistency, we restrict ϕ to small values ($\phi << 0.5$). The minimum contract value under ϕ states that only coalitions with higher contract values ($P_S^{CRCT} \geq P_\phi^{MIN}$) will be accepted.

At this point, if a coalition S whishes to join the market, it has to choose a contract value that fulfill the two conditions. For simplification, we consider that coalitions will always apply the same economically consistent strategy of announcing the highest possible contract value that obeys the reliability rule (a value we denote by $P_\phi(S)$). If $P_\phi(S) \geq P_\phi^{MIN}$, meaning that it also obeys the minimum contract value rule, then S annonces this value on the market : $P_S^{CRCT} = P_\phi(S)$. Otherwise, coalition S is not able to enter. Basically, a coalition

S is valid if and only if :

$$\begin{cases} \forall t,\ Pr[P_S(t) \leq P_\phi(S)] \leq \phi \text{ (reliability rule)} \\ and\ P_\phi(S) \geq P_\phi^{MIN} \text{ (min value rule)} \end{cases}$$

We now choose a very simple utility function 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 indicates that we favorise small coalitions, mainly because they are easier to maintain in term of communication.

$$\mathcal{U}_{\phi,\ P_\phi^{MIN}}(S) = \mathbf{1}_{S\ valid} \frac{P_\phi(S)}{|S|}$$

Obviously, maximising this utility function amounts to maximizing the coalition contract value with the minimum possible number of agents.

In order to illustrate what is done in the following, lets consider a very simple example with two agents, say i and j, with gaussian value probability distributions $\mathcal{P}_i = \mathcal{N}(\mu_i, \sigma_i)$ and $\mathcal{P}_j = \mathcal{N}(\mu_j, \sigma_j)$, such that the joint probability distribution $\mathcal{P}_{ij}$ of the coalition $\{ij\}$ is also a gaussian with the following parameters :

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

with ρ_{ij} the Pearson correlation coefficient between P_i and P_j. We can easily write the reliability condition $Pr[P_{ij}(t) \leq P_\phi(ij)] \leq \phi$ as :

$$\frac{1}{2}\left[1 + erf\left(\frac{P_\phi(ij) - \mu_{ij}}{\sigma_{ij}\sqrt{2}}\right)\right] \leq \phi$$

The strategy of $\{ij\}$ consists in maximizing its contract values as long as it respects this inequality, e.g to annonce $P_\phi(ij) = \mu_{ij} - \sigma_{ij}\sqrt{2}erf^{-1}(1-2\phi)$. It thus appears (as it was intuitively understandable) that, for equivalent sizes, coalitions with low relative standard deviations (σ_{ij}/μ_{ij}) are able to announce higher contract values.

What this paper investigates in the following is the developement of an algorithm that organizes prosumers such that the synergy term of the standard deviation ($\rho_{ij}\sigma_i\sigma_j$ in the example above) is minimized. In such settings, and through the clique percolation procedure (section 4), we will show that coalitions with low relative standard deviation, e.g high utility coalitions, can be computed.

IV. Correlation Graphs

Applying this technique to the production timeseries of our prosumers (after seasonal effects had been removed) leads to well defined clusters. Indeed, depending on where they are (work, home...), groups of people consume differently but similarly with others of the same group. In the same maner, groups of DER that are subject to similar meteorological conditions produce accordingly. Nevertheless, these clusters of strongly correlated prosumers are the exact opposite of what we are seeking. We thus opt for reversing the metric ($d_{ij} = \rho_{ij}^2$) such that "strong links" become links between uncorrelated timeseries and "weak links", between (anti)correlated timeseries. As expected, independently of the ϵ parameter, the graph exhibits henceforth much less clustering and communities are hardly visible. Therefore, using classical clustering or community detection algorithms do not provide very good results (little better than a random algorithm from the utility point of view). However, as it is intuitively understandable, cliques of this graph exhibit very good utility values. But, as some of them may be too small to reach the grid minimum value requirement or because some of them may benefit from additional agents even if they don't form a clique anymore, we used a slightly modified clique percolation algorithm in order to expand them while still controling their expansion. Next section provide some details on clique percolation.

V. Clique Percolation

For simplicity, in this paper, we wish to keep the coalitions separated and leave the management of overlapping coalitions for future work. We thus implemented a simple heuristic that consider 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 :

$$\tau_i(S) = \frac{\mathcal{U}_{\phi,\ P_\phi^{MIN}}(S) - \mathcal{U}_{\phi,\ P_\phi^{MIN}}(S - \{i\})}{\mathcal{U}_{\phi,\ P_\phi^{MIN}}(S)}$$

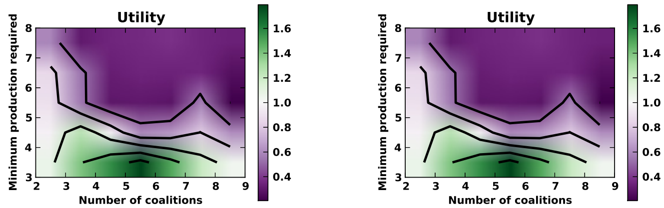
If node i belongs to multiple coalitions, the only coalition retaining node i is the one that maximizes τ_i.

We now test our implementation of the clique percolation against well-known benchmark networks in community detection, namely the Zachary karate club, and the UK faculty network. Results can be seen in figure 1.

Left graph is the Zachary karate club graph, shapes represents real communities and color the algorithm output. We can see that only one node was misclassified. Right graph is the UK faculty graph where the algorithm is able to split the graph in three distinct communities.

VI. Results

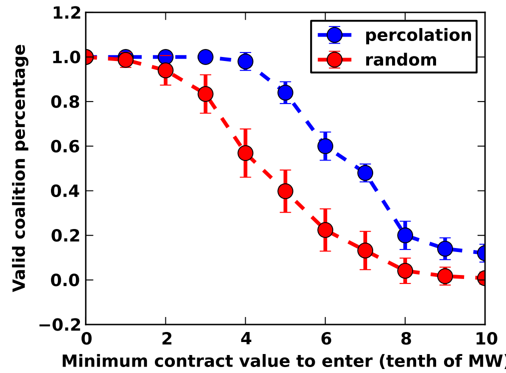
We now test the algorithm on a 200 prosumers example. Simulations were run from february 2006 to december 2010 such that we are dealing with 200 hourly sampled timeseries of available power over this period. The first step consists in removing seasonal effects attributable to wind, sun, and temperature natural patterns. We then construct the complete (un)correlation graph as explained in section 3.



At this point, we have four degrees of freedom, namely the reliability (ϕ), the required power to enter the market (P_{ϕ}^{MIN}), the prunning parameter (ϵ) in order to obtain the ϵ-graph, and the number of desired coalitions (N_{COAL}). A first remark would be that ϵ has a strong impact on the number of possible coalitions as it is directly responsible for the number and sizes of the cliques in the ϵ-graph. If ϵ is too low, the ϵ-graph will not provide enough cliques, conversely, if ϵ is too high, we loose important information and cliques are able to percolate without topological constraints, leading to completely overlapping giant coalitions. Besides, high values on ϵ give rise to higher computational time as more complete graphs are considered. For $N_{COAL} = n$, we thus choose ϵ as the smallest threshold such that the resulting graph contains at least n non overlaping cliques.

The first two parameters (ϕ and P_{ϕ}^{MIN}) shape the utility function such that, if ϕ is close to zero, the reliability requirement is very high and only small values of P_{ϕ}^{MIN} could lead to valid coalitions (and positive utility values). Conversely, the higher ϕ, the less constraints are imposed to the coalitions and valid ones can arise for a larger spectrum of P_{ϕ}^{MIN} (see figure 1, left plot). In the following, we fix the reliability to a given empirical value and observe how the coalitions evolves for different values of P_{ϕ}^{MIN}.

As is visible on the right plot of figure 1, when grid constraints are fixed and only the number of coalitions varies, we can see an increase in social welfare (sum of coalitions utilities) up to a maximum point before a decrease. The reason is that increasing the number of coalitions allow, with our algorithm, more coalitions to be stable and enter the market, but there is a point where nodes bringing stability are not sufficient in coalitions to make them pass the grid requirements, and some coalitions start to fail with zero utility. Moreover, reckon that increasing N_{COAL} means also increasing ϵ, leading to graphs where information is flooded, meaning that the algorithm performances decrease. Naturally, it is also visible that small values of P_{ϕ}^{MIN} lead to higher utilities because coalitions are able to announce higher contract values.



Now that the behavior of the utility is more clear, we have a look at how the algorithm performs. As a comparison, we use a completely random algorithm that only ask for a number of coalitions and partition the agents in a random fashion. For consistency, we always average the results of this algorithm over 100 realization and the errorbars in the plots stands for the standard deviations of the results.

VII. Conclusion

The conclusion goes here.

Acknowledgment

The authors would like to thank...

References

[1] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.