Vincent Gauthier
Institut Mines Telecom
Nano Innov, Saclay
France

November 19, 2015

Dr. Jianhui Wang
Editor-in-Chief
Journal Transactions on Smart Grid

Dear Dr. Jianhui Wang,

The authors would like to thank you for giving us the opportunity to revise and resubmit this manuscript. We resubmit this revision before the agreed upon deadline, December 21, 2015. We appreciate the time and details provided by each reviewer and by you and have incorporated the suggested changes into the manuscript to the best of our ability. The manuscript has certainly benefited from these insightful revision suggestions.

See bellow our replies to the reviewers comments:

Reviewer: 1

Comment 1: However I can see a case study is presented and the results are summarized in the last section, but I recommend you to give more detail about the case study conditions. I can see so many details are presented in the main body of the manuscript but it is difficult to track till end of the paper.

Comment 2: I do recommend you also to have a general schematic of the system to make it clear the elements of your system.

[**Answer to 1 and 2**] We agree with the reviewer that a general schematic would help the reader. In order to illustrate the overall process we added a figure (Figure 1) in the article. This figure explains the main steps from the simulation of the agents production time series to the formation of the coalitions. We also added a table of the main notations that might help the reader.

Comment 3: In fig. 5, how does the sampling frequency affect the volatility? How does the sampling frequency of your data affect the variance? There are some paper that demonstrates the effectiveness of the available resources data resolution on the expected variability of the renewable energy systems, I think this could be interesting if you can provide some information in this regard.

[**Answer**] The results presented in this paper come from hourly sampled weather data. Therefore, all the correlations between the agents are derived from hourly sampled time series. This low frequency sampling has the effect of hiding short time intermittencies that are present in both wind speed and solar irradiance real dynamics. According to [1], extreme events in the solar irradiance and wind power fluctuations are much more frequent in short time scales than expected with a Gaussian process. This is especially true and understandable for solar irradiance that can change drastically in case of clouds passing in a sunny sky for instance. Although we did not, in this paper, study how the correlation structure evolves with the sampling frequency of the data, we believe that the variance of the power distribution of the coalitions will increase with the sampling frequency. Nevertheless, as our algorithm tries to minimize intra-coalition correlations, we believe that short time fluctuations will tend to impact less our coalitions than the ones obtained with the other presented strategies.

We added the following text at the end of section 3 :

Note that this low sampling frequency might hide short time intermittencies (see [1] for more information). Studying the evolution of the correlation structure with the sampling frequency is out of the scope of this paper, but might lead to interresting results.

Comment 4: What is the affect of using stochastic modeling of the resources on the coalitions resulting from the proposed method? Using a real data set is a good case study, but in the conclusion you mentioned "the coalitions resulting from our algorithm better withstand losses of agents", did you do any comparison? It is not clear how you came up to this conclusion. I think you should have an apple-to-apple comparison, stochastic modeling can help you.

Comment 5: There are several parameters that makes the systems fail, in page 7 what do you mean by node failure? Do you mean lack of resources OR failure of power electronics interfaces of renewable energy systems? Please clarify that in the paper.

[**Answer to 4 and 5**] Lack of resources is a key issue of this paper. However in figure 8, we specifically concentrate on the power electronics or line failures using a simplified algorithm. More precisely, we want to study how coalitions resulting from different formation algorithms respond to failures. Failures are considered random and when some agent fails, it is deleted from its coalition. This has consequences on the production of the coalition that may cause the violation of one of the market rules (stability and minimum production), which in turn will lead to the deletion of the coalition from the market. We used this method in order to compare how the coalition structures obtained with the three different algorithms (correlated, de-correlated, and random) respond to this type of failure. We concluded that the coalitions resulting from our de-correlation algorithm were able to better sustain agents loses than the two others. We agree with the reviewer and added, for clarity, a section in the appendix (see section Resilience algorithm) that gives the pseudo code of the process discussed above (reproduced here as algorithm 1).

We also added the following lines in the article:

Data: $CS = \{S_1, S_2, ..., S_k\}$ Coalition structure ;
$pool = \cup_{k \in N_{COAL}} S_k$;
while $pool \neq \emptyset$ **do**
  Select randomly agent i from pool;
  i fails $\Rightarrow$ Remove i from its coalition : $S_k \leftarrow S_k - \{i\}$;
  $pool = pool - \{i\}$;
  if $Pr\left[P_{S_k} < P^{MIN}\right] > \phi$ **then**
    S_k fails $\Rightarrow$ Remove S_k from the market
  end
end

Algorithm 1: Random failures algorithm

We consider here the case of random failures of the power electronics of some agents that has the consequence of preventing them from participating in the market. Therefore, the notion of resilience we will use in the following can be seen as the ability of the coalition structures to inject stable power in the grid when some of its internal agents are removed.

Comment 6: Fig. 5 is a great visualization of the problem, but how are you going to come-up to a final decision? Can you provide formulation to make decision using decision theory? I think this will help and support your great work.

[**Answer**] We thank the reviewer for this comment. In figure 5 we represented the distribution of the coalition structures using Monte Carlo sampling as comparison between the results of the three algorithms. All algorithms need a desired number of coalitions that they should output. In our work, the decision on the number of coalitions is therefore assumed to come from the user, and is not investigated. Given the desired number of coalitions, the algorithms use the utility function (eq. 11) as the decision criterion between the different coalition structures (all containing the same number of coalitions). This can be seen in more details in the pseudo-code of the algorithms in the appendix (see section Algorithms of the appendix). We also reproduce the pseudo code of the greedy optimization below (alg. 2). Our algorithm for instance starts with seeds in the correlation graph and expands them as long as the utility is improved. It stops when no agent is able to increase the coalitions utilities.

Comment 7: In general the paper is in well format, however it needs some grammar revision as well as proof reading, I saw couple of errors such as the question marks in page 3, column 2.

[**Answer**] We tried to correct all the remaining errors in the paper.

Reviewer: 2

Comment 1: Using a Table with ALL notations at the very beginning of the manuscript.

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

Algorithm 2: Local greedy optimization algorithm

[**Answer**] We respond to the reviewers comment and added a table of the main notations in the article (figure 2). Because of space limitation, we were not able to put all the notations in this figure. Therefore, we also added a table of all the notations in the appendix.

Comment 2: Making clear the fact which grid/system operator is referred to (distribution vs transmission)

[**Answer**] In this paper, we consider aggregations of prosumers located in the distribution network as stated in the introduction: Since production could be located down to the very end of the distribution networks, nodes that were simply pure loads yesterday could behave tomorrow as generators or loads, further complicating energy usages. In this paper, we focus on these "nodes" in the distribution network that can produce and consume electricity.

For clarity, we added the following line in the beginning of section 4:

We consider a set A of N prosumers of the distribution network.

Reviewer: 3

Comment 1: In Section III, is load uncertainty included in the generation of prosumer patterns? If yes, what load patterns were used? What happens if the load is so high that prosumers cannot generate energy?

[**Answer**] In this paper, we generate consumption historical series using real temperature traces (see section Consumption of the appendix). More precisely, we decompose the consumption of a prosumer into a heating and an electronic consumption term.

$$P_i^D(t) = \mathcal{F}_i^{heat}(\tau(t), t) + \mathcal{F}_i^{elec}(t) \tag{1}$$

where $\mathcal{F}_i^{heat}(\tau(t), t)$ is the power curve that maps the temperature $\tau(t)$ to a heating consumption, and $\mathcal{F}_i^{elec}(t)$ computes the consumption of agent i (other than heating) at a given hour of the day. The purpose of this decomposition is to reproduce both seasonal patterns as well as daily patterns of consumption in a simple model.

In the simulation, all agents have a desired inside temperature T_i, supposed to be a constant for simplification. By using thermodynamic laws $\mathcal{F}_i^{heat}(\tau(t), t)$ can be approximated by :

$$\mathcal{F}_i^{heat}(\tau(t), t) = \frac{B_i}{R_i}\left[T_i - \tau(t)\right] \tag{2}$$

where B_i is the surface of thermal exchanges for agent i and R_i is their thermal resistance.

We further denote by Ω_i the maximum consumption possible for agent i, which is basically the sum of all its appliances powers. We also denote by $\omega_i(t) = \{\omega_i(t_0), ..., \omega_i(t_{24})\}$ the vector of the average fraction of Ω_i used for each hour $(\forall i,\ \forall t,\ \omega_i(t) \in [0, 1])$. We can therefore write :

$$\mathcal{F}_i^{elec}(t) = \Omega_i(\omega_i(t) + \epsilon) \tag{3}$$

where ϵ is a noise term drawn from a normal distribution. The vector $\omega_i(t)$ enables us to easily differentiate agent consumption behaviors. Business or residential areas for instance can be easily distinguished with this kind of model.

At time t, if the consumption of a prosumer is larger than its production, then this prosumer will consume energy from its coalition rather than contributes to the production. If all prosumers of a same coalition are strongly correlated and tend to behave in the same way this results in a very unstable output production. However, in the case of uncorrelated prosumers, over-production of some of them might tend to compensate the under-production of the others, yielding a more stable output. In the case of a prosumer that almost never produce, its marginal utility would be very small or negative, such that it will not be selected in the expansion phase.

Comment 2: The selection of the parameter alpha of the coalition utility function [eq. (14)] relies heavily on the Gaussianity assumption for the generated net power, which may not hold.

[**Answer**] We agree with the reviewers comment. The purpose of the alpha parameter is to control to some extent the expected sizes of the coalitions.

In order to obtain an analytical formulation for this parameter, we used the assumption that the distributions of the coalitions production were Gaussians. When the number of agents in the coalitions increases, the distribution of the coalition will tend to a normal distribution for independent agents. Although independence is generally not true and is not required for our method, we found that using this assumption for tuning alpha gives good results in practice. Other parameter selection methods such as grid search could also be used although not presented in the article.

Comment 3: The selection of the threshold epsilon for the decorrelation graph [eq. (15)] appears to incur exponential complexity, as it requires availability of cliques of the graph.

[**Answer**] We agree with the reviewers point that finding cliques in graphs is usually a hard problem. In order to overcome this difficulty when the number of nodes is not small, we restrict the algorithm to triangles rather than cliques. Finding $\epsilon\star$ through (eq.(15)) with $k = 3$ (looking only at triangles) requires to count the number of triangles in a graph. There exist several techniques to achieve this (see [2]). One of the most straightforward is:

$$N_{triangles} = \frac{1}{6} Tr \left[A^3 \right] \tag{4}$$

Where A is the graph adjacency matrix. This leads to a counting algorithm which runs in $O(n^q)$ where q is the matrix multiplication exponent (considered to be less than 2.4).

For finding the starting seeds for the clique expansion, we need not only to count the number of triangles, but we also have to enumerate them. Again, several techniques exist such as node-iterator or edge-iterator (more information in [2]).

We clarify this point by adding the following text in section IV-A:

In practice, since finding cliques requires exponential time, we use triangles [14] (k = 3 in eq. 15) rather than cliques as soon as the number of nodes is not small.

Comment 4: Implementation of the algorithm also requires the correlation coefficient between any two prosumers, which may as well not be available.

[**Answer**] We agree with the reviewers comment. In order for our algorithm to work, we need the correlation coefficients between any two prosumers that can be obtained from historical time series. Having these series is an assumption of our work.

Reviewer: 4

Comment 1: The reviewer is wondering if the research question is application for the general context or only for specific environment. In general, it would be expected that the prosumers have their own freedom to choose their aggregator, similar way to have the electricity contract with the energy supplier. Forcing the end users to be in dedicated coalition would be realistic.

[**Answer**] In this paper, we indeed do not consider agent freedom to choose what coalition to join. We only consider the maximization of a global utility function that seeks de-correlation inside the coalitions. Perhaps, a more realistic framework based on Game theory will consider a utility function per agent with different parameters. Each (rational) agent would then focus on the maximization of its own profits, and coalition structures might emerge from series of split and merge actions. Although not considered in the present work, we believe that building a game which would take into account the stability of power injections through the correlation structure might be possible.

Comment 2: The reviewer was not convincing about the statement in page 4, line 21-24, If S always procedures a higher contract value. The reviewer believes that if S deviates from the scheduled value, even if higher extra-production it should still have the penalty cost.

[**Answer**] We agree with the reviewer that over-producing is also problematic toward the grids stability. And coalitions should pay penalties if deviations (positive or negative) from the contract occur. However, in our framework, we make the distinction between the available power of a coalition, and the power it really injects in the grid. Although the latter has to be equal to the contract, we allow the former to be larger since the difference can be used for storage or other usages. We changed the misleading statement to:

If S has an available production always greater than P_S^{CRCT}, it is losing some gains since it could have announced a higher contract value.

Comment 3: The reviewer is doubting about the interaction model developed in the paper, specifically mentioned in the second column of page 4. Why should the grid operator concern about the coalition of prosumers with its probability of under-producing? Normally saying, under-producing should not cause any grid issue while significant over-producing from the end users would might cause current congestions or voltage violation for the distribution network.

[**Answer**] We agree with the reviewer that, in the case of a large penetration of renewables in the distribution network, the production capacity should be larger than for traditional power plants. This is the consequence of the uncertainty that comes with most renewables. It results from this point that, at certain periods, the production might be drastically larger than the consumption, which might indeed cause various problems in the network. Storage, demand-side managment, inter-connection with neighboring countries, as well as energy markets are often assumed to provide ways of stabilizing the grid in such situations.

Nevertheless, we want to point out that, in this paper, we focus on the market side of the system rather than the grid pure stability. More precisely, we consider the formation of coalitions of prosumers that seek to sell energy to some entity (grid operator, aggregators). If this entity buys a contract, the coalition commits to injecting exactly P_S^{CRCT} :

$$\forall t,\ P_S^{injected}(t) = P_S^{CRCT} \tag{5}$$

Obviously, if the actual injected power deviates from the contract, the coalition should pay penalties proportional to $\sum_t \left| P_S^{injected}(t) - P_S^{CRCT} \right|$. However, within this framework, we build the coalitions by looking at their available power $P_S(t)$, i.e, the maximum production that a coalition can inject at time t : $P_S^{injected}(t) \leq P_S(t)$. In this situation, we believe that the risk $Pr[P_S(t) < P^{CRCT}]$ is a key quantity since, in this case, the contract cannot be respected. In cases where the available production is higher than the contract $P_S(t) > P^{CRCT}$, it might be possible to shut some renewables down, or to use storage, to maintain the power injection at the contract value.

We added the following text in the introduction as to clarify this point :

We rely on a market representation of the interactions. Coalitions of prosumers estimate a production capacity for an upcoming period and announce it on the market. If some other entity buys the contract, the coalition has to inject exactly the contracted amount of power at any time. Obviously, deviations will cause financial penalties to the coalitions, such that respecting the contracts is paramount. Since prosumers use exclusively renewables, we are particularly concerned with the probability that the available capacity of a coalition falls below its contract value. In this case, storage or demand side managment might be solutions to avoid penalties, but this probability should still be maintained low.

Comment 4: Keywords are missing in this paper!

[**Answer**] We have considered the reviewers request and added keywords at the beginning of the article.

Comment 5: The introduction part is too long with well-known context that can be found in many previous works. It is recommended to refer more to the literature and shorten the introduction part to go direct to the research problem/questions

[**Answer**] We agree with the reviewers comment and we shortened the introduction.

Comment 6: Terminologies of reliability and stability have been used similarly here. Actually, power system researcher would consider stability with different meaning.

[**Answer**] In this paper, we used the word stability to characterize the smootheness of the power output of the coalitions. In this paper, this smootheness appears as a consequence of the variance of the coalitions power distributions. Therefore the stability and reliability concepts of our paper are strongly connected but we wanted to make the distinction. Naturally, we agree with the reviewer that these terms are used here in a special way, and might be considered differently by power system researchers.

[1] M. Anvari, G. Lohmann, M. Reza Rahimi Tabar, M. Wächter, P. Milan, D. Heinemann, Joachim Peinke, and E. Lorenz. Suppressing the non-Gaussian statistics of Renewable Power from Wind and Solar. 2015.

[2] Thomas Schank and Dorothea Wagner. Finding, Counting and Listing all Triangles in Large Graphs, An Experimental Study. *Framework*, 001907:3–6, 2001.