Smart grid demand response strategies for datacenters

[1] *J. Muraña <jmurana@fing.edu.uy>* [1,2]*, S. Nesmachnow <sergion@fing.edu.uy>*[1]
S. Iturriaga <siturria@fing.edu.uy>[1]*, S. Montes de Oca <smontes@fing.edu.uy>*[1]
G. Belcredi <gbelcredi@fing.edu.uy>[1]*, P. Monzón <monzon@fing.edu.uy>*[2]
V. Shepelev <v.shepelevvd@susu.ru>[2,3]*, A. Tchernykh <chernykh@cicese.mx>*

[1] *Universidad de la República, Uruguay*
[2] *South Ural State University, Russia*
[3] *CICESE Research Center, Ensenada, Mexico*

Abstract. This article presents demand response techniques for the participation of datacenters in smart electricity markets under the smart grid paradigm. The proposed approach includes a datacenter model based on empirical information to determine the power consumption of CPU-intensive and memory-intensive tasks. A negotiation approach between the datacenter and clients and a heuristic planning method for energy reduction optimization are proposed. The experimental evaluation is performed over realistic problem instances modeling different types of clients. Results indicate that the proposed approach is effective to provide appropriate demand response actions according to monetary incentives.

Keywords: smart grid, computational intelligence, demand response

1. Introduction

Smart grids are the current state-of-the-art technology for electricity networks. They include operation and management features to improve the controlling of production and distribution of energy [1].

Within the smart grid paradigm, a large consumer with flexible power utilization can participate in the electricity market. This is one of the main ideas behind the implementation of strategies oriented to modern smart electric networks, where consumers are associated to the roles of both active clients and market agents [1]. As an active client, a consumer can adapt his electricity demand to peak hours, e.g., by reducing power consumption in peak periods and contributing to flattening the demand curve of the whole electrical system. As a market agent, a consumer can participate in the electricity market and receive an income by providing different services (e.g., by establishing bilateral agreements with an electricity generation company or by participating in periodic auctions for smart grid management).

Within the smart grid paradigm, demand response planning strategies are needed to manage energy consumption and be able to participate in the market, in different roles. Specific techniques are needed to plan those activities that consume energy, i.e. by advancing or deferring their execution. In addition, the impact on global energy efficiency, and the possible degradation of QoS offered to users must be analyzed. These planning techniques are essential to ensure the correct use of energy resources and to guarantee the energy efficiency of large flexible consumers.

This article describes a proposal for demand response strategies on datacenters, allowing them to participate in the electric market and provide ancillary services. Datacenter can adjust power consumption to help the electric network to fulfill specific goals: they are able to consume available surplus of energy by executing complex tasks that demand large execution times, or they can defer activities (i.e., tasks execution) in periods where energy is more expensive and/or power generation is lower than normal. Furthermore, their thermal/cooling infrastructures demand significant energy consumption and provide a large inertia. Thus, they can be used to interact with a smart electric grid.

In this line of work, the research reported in this article is based on a negotiation using a pricing mechanism that guarantees that datacenter operators can extract load shedding from tenants. The proposed strategy for demand response allows implementing a smart management of the electric grid, achieving a rational utilization of energy sources, and the correct use of information technologies to improve decision-making processes within modern smart grids.

2. Datacenter Scheduling Problem

Given a reduction request from the electric market, the optimization problem consists in minimizing the total monetary incentive rewarded to tenants and the cost of using the on-site generator in order to meet the reduction target. The energy reduction must be maintained for a time horizon T. The formulation is as follows:

- A set of discrete timesteps t in $[0,T]$.
- A target reduction βrequested by the electric market.
- A set of tenants (or clients), $C=\{c_1,\ldots,c_{¿C\vee¿\}¿}$.
- A workload of tasks for each tenant c_j, $W_j=\left\{w_j^1,\ldots,w_j^{|W_j|}\right\}$.
- Let $DF_j^i=1$ if task w_j^i is deferrable and $DF_j^i=0$ if it is non-deferrable.
- Let DD_j^i be the due date of task w_j^i.
- Let MP_j^i be the monetary penalty received by tenant c_j if the due date of task w_j^i is not met.
- Let RI be the monetary incentive for each tenant for each energy unit reduced.
- A workload schedule for each tenant c_j with no monetary incentives (i.e. $RI=0$), $WS_j=\{ws_1,\ldots,ws_{¿C\vee¿\}¿}$.
- Let DP_j^tbe the power requirement at each time t of each workload schedule ws_j.
- Let FT_j^i be the finishing time of task w_j^i for schedule ws_j.

- Let $VD_j^i = 0$ if the $FT_j^i \leq DD_j^i$ for schedule ws_j, otherwise $VD_j^i = 1$.
- Let define the total monetary penalty of a schedule ws_j for a tenant c_j as $Y_j = \sum_{i=1\ldots|W|} VD_j^i \times P_j^i$.
- Let the function γ_j determine the new schedule $\overline{ws_j}$ with a power requirement $\overline{DP_j^t}$ for tenant c_j given incentive RI, $\gamma_j(RI) = \overline{ws_j}$.
- Let define the energy reduction function between $\overline{ws_j}$ with respect to the schedule ws_j as $\delta(\overline{ws_j}) = min\{\overline{DP_j^t} - DP_j^t, t \in T\}$.
- Let GP^t be the energy generated at each time t using the on-site generator.
- Let α be the monetary cost per unit of energy of using the on-site generator.

The objective function defined by Eq. 3.1.

$$min\, z = \sum_{j=1}^{|C|} \delta(\gamma_j(RI)) \times RI + \sum_{t=1}^{T} GP^t \times \alpha \tag{3.1a}$$

subject to

$$\beta \leq \delta(\gamma_j(RI)) + GP^t \tag{3.1b}$$

$$z \leq \sum_{t=1}^{T} \overline{DP_j^t} \times \alpha \tag{3.1c}$$

- The objective (3.1a) is to minimize cost for the datacenter operator (i.e., the money paid to tenants plus the cost of using the on-site generator) in order to meet the reduction target. Constraint (3.1b) states the total energy reduction must be at least β per timestep. Finally, constraint (3.1c) indicates that the total monetary equation must be less than the cost of powering the whole datacenter using the on-site generator alone.

3. Demand Response Approach

Datacenter operation. The grid operator offers a monetary incentive to the datacenter administrator for reducing a certain amount of energy from its consumption during a certain time. To achieve the necessary reduction, the next market mechanism was implemented.

In the proposed market mechanism, the operator can induce a reduction on client's power consumption diminishing the need of brown energy, using a parameterized supply function represented in Eq. 4.1, where r_i is the power reduction for client i, D is data center's power reduction target, b_i is the client offer for reducing the power consumption by r_i and p is the market clearing price determined by the operator [2].

$$r_i(b_i, p) = D - \frac{b_i}{p} \quad (4.1)$$

The market mechanism for reducing D amount of energy is exercised in four steps in an iterative approach:

(i) The datacenter broadcasts the supply function to the clients, $r_i(b_i, p)$.

(ii) Each client i bids a reward b_i for reducing power r_i, to maximize its utility.

(iii) The datacenter determines the market clearing price p and the energy to produce on-site y (with generation cost α) by minimizing the total cost.

$$p(b_i, y) = \frac{\sum_i b_i}{(N-1)D + y} \quad (4.2)$$

$$y = argmin\,(D-y)\,p + \alpha y, 0 \leq y \leq D \quad (4.3)$$

The first-order optimality condition for Eq. 4.3 gives the value for y:

$$y = \sqrt{\frac{\left(\sum_{i=1}^{N} b_i\right) ND}{\alpha}} - (N-1)\,D \quad (4.4)$$

(iv) If p and y converge, latest bids are accepted and energy reduction is scheduled by the clients, else the operator broadcast the new supply function with the updated value for p.

The strategy used by the datacenter solves the location problem based on a proximal method [3]. A distributed solution is generated for each agent. In the algorithm, D is the power reduction target, *price* is the market clearing price per Watt, N is the number of tenants and j is the tenant id. Function *client_evaluation(price,j)* corresponds to the offer evaluation of the tenant j, considering its SLAs. This function returns the energy reduction committed by the tenant (*reduction[j]*), according to the price, *bid*[*j*] is the offer of tenant *j* for reducing the power consumption, y_k is the iteration variable, which at the end of the negotiation corresponds to the power generated by the on-site generator. The cost of generate one Watt using the generator is denoted α. The parameter ϵ is a measure of the compliance of the coupling restriction.

Algorithm 1 Datacenter market mechanism

INPUT: D (power reduction target), $price_0$

OUTPUT: $price, on-site\ generation$

$k \leftarrow 0$ ¿iteration step

$price_k \leftarrow price_0$

while $\epsilon \geq \epsilon_{min}$ **do**

  for $j=1$ to N **do**

    $reduction[j] \leftarrow client_{evaluation}(price_k, j)$

    $bid[j] \leftarrow (D - reduction[j]) \times price_k$

  end for

  $y_k \leftarrow max\left(\sqrt{\left(\sum_j bid\right) \cdot \frac{ND}{\alpha}} - (N-1)D, 0\right)$

  $price_k \leftarrow \sum_j bid[j] / ((N-1)D + y_k)$

  $\epsilon \leftarrow \left\| \left(y_k + \sum_j reduction[j] - D\right) / D \right\|$

  $k \leftarrow k+1$

end while

$on-site_{generation} \leftarrow y_k$

Client offer evaluation. To evaluate the monetary offer of the datacenter administrator and determine the amount of power to be reduced, clients simulate the execution of their workload, applying an energy optimization strategy. The monetary offer of the datacenter administrator is accepted if the net income obtained from the energy reduction minus the loss the client must pay in case of not complying with the SLA with his users, is greater than zero. In any case, different trade-offs are obtained for different monetary offers from the negotiation. These trade-offs can be considered in case the datacenter cannot meet the desired power consumption reduction, to account for different compromises between the problem objectives (energy reduction and cost).

The energy optimization strategy proposed in this article aims at maximizing the profit of clients by reducing the active cores of the servers, thus lowering the energy consumption according to the offer received from the datacenter administration. A heuristic procedure is applied: Active Cores Reduction (ACR) whose main idea is to select the best scheduling according to its profit, considering all combinations of active cores. The details of the ACR heuristic are presented in Algorithm 2, where *price* is the offer per reduced Watt and *reduction* are the watts that the client is willing to reduce according to the offer. Function *schedule* simulates the execution of the workload considering *cores_number* active cores, out of the *server_cores* total cores available in the client. In turn, function *eval* evaluates the profit and the reduction of the solution scheduling *sol*.

Algorithm 2 Energy optimization strategy

INPUT: $price$
OUTPUT: reduction
$profit \leftarrow 0$
$reduction \leftarrow 0$
$cores \leftarrow server_{cores} \times serves_{numbers}$
for $cores_{numbers}$ in $cores$ **do**
  $sol \leftarrow schedule(cores_{number})$
  $reduction_{ax}, profit_{ax} \leftarrow (price, sol)|$
  if $profit_{ax} > profit$ **then**
    $reduction \leftarrow reduction_{ax}$
    $profit \leftarrow profit_{ax}$
  end if
end for

Client scheduling simulation. Clients are providers of HPC services to single users. Batch tasks arrive to the system and they are queued until a server has the capacity to execute it (i.e., it meets the task requirements, such as available cores, memory, and estimated execution time).

A simulation-driven approach is applied to determine the cost of implementing a certain energy optimization strategy. An ad-hoc simulator is used, due to the limited capabilities of existing datacenter/cloud simulators to provide an accurate environment for implementing the main features of the proposed approach.

The simulation period is divided into intervals of equal duration $\int ¿_d ¿$. At each interval, the scheduler assigns the arrived tasks to the servers, taking into account the current capacity of each server and the scheduling strategy criteria. The number of intervals that a task is running on a server is calculated as $ct/¿t_d+1$, where ct is the completion time of a task (in seconds) and $\int ¿_d ¿$ is the duration of each interval (in seconds). The completion time of a task is defined as its size in millions of instructions per second (MIPS) divided by the assigned core speed (also, in MIPS).

Power consumption model. In order to estimate the power consumption of a scheduling, the novel quadratic power consumption model, developed using real experimental data, was used [4].

4. Experimental Analysis

Specific instances were generated to evaluate and validate the proposed model for datacenter participation in the electricity market. Preliminary results are reported for instances with 5 clients, 5 servers and 1500 tasks each. Further experimental evaluation is reported in [2]. Regarding the computational infrastructure, servers with 24 cores are considered, with Intel Xeon CPUs and a processing speed of 3000 MIPS. Experiments were developed in Java SE 1.8, and executed on National Supercomputing Center, Uruguay [5].

Fig. 5 shows the result of the power reduction negotiation between the datacenter administrator and the clients, for small instances. The offer (per Watt) to the clients is on the independent axis. Blue squares are the evolution of the clients reduction (CR) through the negotiation, red circles are the values of y_k, and green triangles are the power generated on-site to cover the power reduction target (OG=D-CR). For the small instances, the negotiation quickly reaches a power reduction of 600 W, with a low monetary cost. However, larger power reductions require more iterations and a greater monetary offer. This slowdown in reduction is because, at this point, clients cannot reduce their power consumption without a significant impact on user performance.

Table 2 reports a summary of the negotiation for small instances. Three negotiation steps are considered: the first offer, an intermediate offer, and the last offer (that is, when the negotiation algorithm ends according to the stopping criteria). Column k is the negotiation step, *price* is the offer in step k, *CR* is the reduction obtained from clients, *OG* is the on-site generation, ϵ evaluates the compliance of the coupling restriction, and *cost* is the monetary value that the datacenter administrator must invest to achieve the target reduction (equation 5.1).

$$cost = CR \times price + OG \times \alpha \tag{5.1}$$

Table 2. Negotiation summary

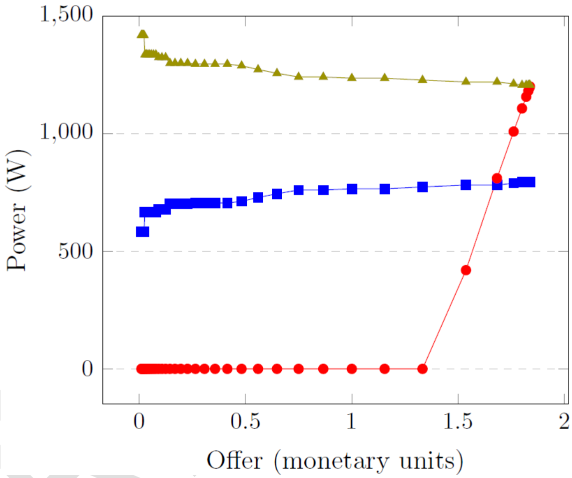
DPEM - small - 1.0						
k	*price*	*CR*	y_k	ϵ	*OG*	*cost*
1	0.012	582	0	70.90	1418	2842
20	0.227	701	0	64.95	1299	2757
41	1.840	794	1200	0.03	1206	3872
DPEM - small - 0.8						
k	*price*	*CR*	y_k	ϵ	*OG*	*cost*
1	0.012	582	0	70.90	1418	2842

21	0.260	705	0	64.75	1295	2773
41	1.760	1206	802	0.04	794	3710

DPEM - small - 0.4

k	*price*	*CR*	y_k	ϵ	*OG*	*cost*
1	0.012	665	0	66.75	1335	2677
21	0.244	728	0	63.60	1272	2721
42	1.576	1894	0	0.06	126	3205

(a) DPEM-small-1.0



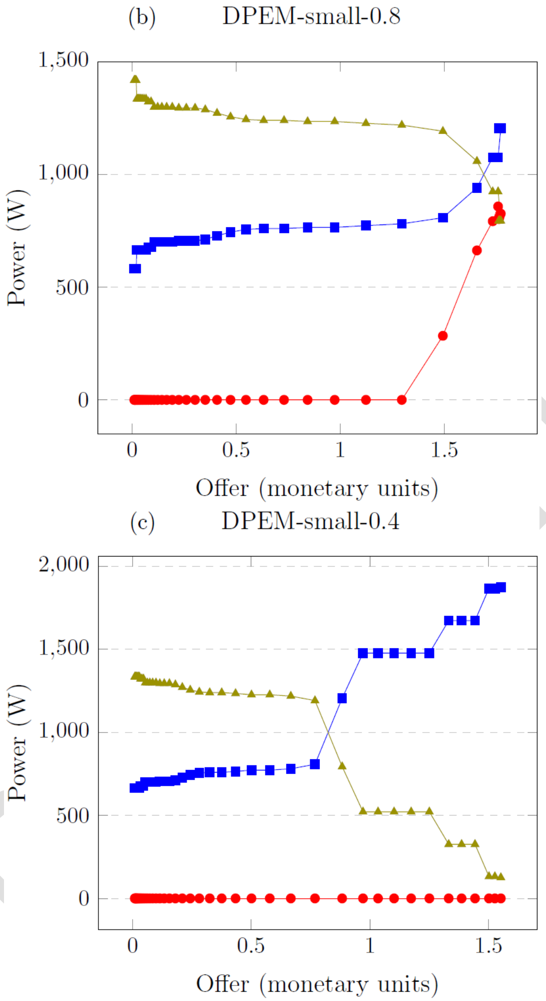


Fig. 5. Negotiation for small instances: red dots - y_k, blue squares - CR, red triangles - on-site generation.

The comparison between instances with different tolerance values shows that when clients are less flexible, the negotiation determines in the last step low values in CR column and high values in OG column. This behavior corresponds to the intuitive idea that in datacenters where clients have less flexible SLAs, the on-site generation is the main option to achieve the target reduction established by the electric provider. Moreover, less flexible instances (i.e. small-1.0) imply large offers.

Results confirm that the proposed negotiation approach is able to properly take advantage of deferring execution tasks to fulfill the requested power consumption reduction.

5. Conclusions

This article studied a negotiation approach for the participation of datacenters and supercomputing facilities in smart electricity markets, an important problem in modern smart grid systems.

A specific case of demand response strategy was studied for colocation datacenters to commit power reductions, according to offers proposed to clients. A decentralized approach was applied for negotiation, where clients do not need to provide strategic information to the datacenter administrator. Instead, each client negotiates a price considering a planning heuristic and the features of the tasks submitted for execution. A model based on empirical information was presented to determine the power consumption of CPU-intensive and memory-intensive tasks, using data from real datacenters.

The negotiation algorithm and a heuristic planning method for energy reduction optimization were experimentally validated over nine realistic problem instances that model different problem dimensions and flexibility of the datacenter clients.

The obtained results indicate that the proposed approach is effective to provide proper demand response actions according to monetary incentives. The system achieved economic benefits for the datacenter operator and for the tenants (by providing rewards for reductions) and for the environment, by reducing diesel use.

Summarizing, clients quickly reached appropriate power reductions, thus limiting the need of using on-site generation by the datacenter. Results confirmed that the problem is inherently multiobjective. Both operation costs and QoS offered to users must be considered in the formulation, and trade-offs between total cost and negotiation offers must be studied. The proposed approach is realistic and efficient, to be implemented in nowadays datacenters and supercomputing facilities.

Acknowledgments. The work is partially supported by Agencia Nacional de Investigación e Innovación (FSE_2017_1_144789). The work of S. Nesmachnow and S. Iturriaga has been partly funded by ANII and PEDECIBA, Uruguay. The authors also want to thank the Centro Nacional de Supercomputación (Cluster.Uy).

References

[1]. J. Momoh. “Smart Grid: Fundamentals of Design and Analysis”. Wiley-IEEE Press, 2012.

[2]. J. Muraña, S. Nesmachnow, S. Iturriaga, S. Montes de Oca, G. Belcredi, P. Monzón, V. Shepelev, A. Tchernykh. “Negotiation approach for the participation of datacenters and supercomputing facilities in smart electricity markets”. In: Programming and Computer Software (2020).

[3]. N. Parikh and S. Boyd. “Proximal Algorithms". In: Foundations and Trends in Optimization 1.3 (2014), 127-239.

[4]. J. Muraña, S. Nesmachnow, F. Armenta, and A. Tchernykh. “Characterization, modeling and scheduling of power consumption of scientific computing applications in multicores". In: Cluster Computing 22.3 (2019), pp. 839-859.
[5]. S. Nesmachnow and S. Iturriaga. “Cluster-UY: scientific HPC in Uruguay". In: International Supercomputing in Mexico. 2019, pp. 1-15.