LOAD MANAGEMENT WITH PREDICTIONS OF SOLAR ENERGY PRODUCTION FOR CLOUD DATA CENTERS

Maurizio Floridia(1), *Demetrio Laganà*(2), *Carlo Mastroianni*(3), *Michela Meo*(1), *Daniela Renga*(1)

(1) Politecnico di Torino, Italy
(2) Eco4Cloud srl, Rende (CS), Italy
(3) ICAR-CNR, Italy

ABSTRACT

Power supply of big infrastructures is today a tremendous operational cost for providers and the expected growth of Internet traffic and services will lead to a further expansion of the computing and networking infrastructures and this, in its turn, raises also concerns in terms of sustainability. In this context, renewable energy generators can help to both reduce costs and alleviate the concerns of sustainability of big infrastructures. In this paper, we consider the case of Data Centers (DCs) composed of a few sites located in different geographical positions and powered with solar energy. Due to the intermittent nature of solar energy, different time zones and price of electricity in different locations, load management strategies are fundamental. We consider predictions of the solar energy production performed through Artificial Neural Networks and we assess the impact of predictions on load management decisions and, ultimately, on the DC performance.

***Index Terms*—** Artificial Neural Networks, Solar Energy, Data Centers.

1. INTRODUCTION

In the last years, the Cloud computing paradigm has emerged as a mean to provide contents, services, computing and storage facilities to end costumers and to companies, in a very dynamic and flexible way. The size and the number of data centers (DCs) have grown accordingly. With the expected further increase of the traffic carried by Internet, pushed by the further diffusion of communication services, by the digitalization of services and products, the trend is not expected to change in the next years.

One of the main challenges coming with the growth of the size and complexity of these big data infrastructures is their energy consumption, estimated to reach 140 billion kilowatt-hours annually by 2020, corresponding to about 50 large power plants [1]. The implications of this huge consumption are manifold and raise a number of concerns. First, the expected increase of energy consumption implies an increase of costs for electricity bills. Second, the size of these infrastructures makes their powering critical per se. Finally, there is a general concern about the sustainability of their growth, concern which is becoming the more and more critical as the general awareness of the impact of energy consumption on climate changes grows. In this context, the introduction of renewable energy sources (RES) appears as an interesting possibility that jointly helps reducing the cost of electricity bills, reduce the burden on the power grid, stressed by the excessive needs of these systems, alleviate the concerns on sustainability and carbon emissions [2]. While RES are very attractive for the above mentioned reasons, their introduction as power supply to DCs requires a careful integration of energy management strategies into the system operation and management.

In this paper, we consider a geographical distributed DC composed of a few sites, each equipped with a hybrid power supply system composed of photovoltaic (PV) panels, energy storage units and access to the power grid. By being in different locations, the sites undergo different electricity prices and, by effect of time zones and different weather conditions, the sites are exposed to different levels of renewable energy (RE) production and grid electricity prices. In order to manage the load among the sites so as to adapt the working conditions of the DC sites, to the price and green energy production, and reduce costs, the prediction of the solar energy production is important. Indeed, similar scenarios were already considered in our previous work [3] under the optimistic assumption that the PV panel production is known in advance. In this paper, we consider the same approach as in [3] but we make the more realistic assumption that the green energy production is not known a-priority but can be predicted. In particular, the predictions are performed using Recurrent Neural Networks (RNN). The results we present in this paper prove that the introduction of load management strategies to reduce the cost and the consumption of geographically distributed DCs is feasible and effective. Marginal difference between the results achieved with a perfect knowledge of the energy production and with predictions can be observed. Other studies in the literature explore the opportunity of energy cost-saving by routing jobs based on electricity prices [4], [5], [6], [7], [8]. However, the introduction of RES and predictions of the

RE production had not been considered in this context.

2. THE SYSTEM

As introduced in the previous section, we consider the multi-site load management strategy proposed in [9] and then studied in [3] in presence of RE generated by PV panels. The strategy is called EcoMultiCloud and is organized in two layers. In the *upper layer* the various DC sites exchange some very basic information, i.e., the value of a few indicators that describe in a very compact way the site working conditions. These indicators are typically as simple as the load, the electricity price, the amount of available green energy, the Power Usage Effectiveness (PUE). This information is exchanged among sites and used to distribute the Virtual Machines (VMs), i.e., the workload, among the sites. In addition, this same information is used to decide possible VM migrations among sites, migrations which aim at adjusting the workload distribution. At the *lower layer*, some strategy that is specialized within each site is used to allocate the workload among servers within the same site.

The scenario that we consider for the derivation of numerical results is as in [3]: four sites compose the geographically distributed DC. They are located in California, Ontario, UK, Germany and have a PUE of 1.56, 1.7, 1.9, 2.1, respectively. The DCs are equipped with a hybrid power supply system that includes, besides the possibility to get power from the grid, PV panels and batteries in which the excess of solar energy can be stored for future usage. We assume a battery charging/discharging efficiency of 85%. Since in our scenario the main goal of the workload management strategy is to reduce the energy cost, the information that the DCs exchange to take workload distribution decisions in the EcoMultiCloud framework is limited to two values: the amount of RE locally available and the price of electricity from the grid in that site. The migration and assignment of some workload is decided based on the following *cost function* computed at each DC i:

$$f^{(i)}_{assign} = [V_i - \max(0, E_i - C_i)]\, P_i \qquad (1)$$

where V_i is the marginal consumption at DC i associated with a VM, i.e., it is the additional consumption in case a VM is assigned to DC i; C_i is the energy consumed for the whole DC i, E_i is the green energy, including both the energy that is expected to be produced by the PV panel in the next time step and what is currently stored in the battery, and P_i is the electricity price expected in the next time step. Basically the term in brackets represents the available green energy in DC i. This term can be 0 in case the consumption is larger than the production, and in this case the cost function is equal to the monetary cost required for buying the electricity needed to run a new VM. Conversely, when green energy is available, the cost function is negative and represents a gain.

The load is assigned to the DC who exhibits the smallest cost function, provided that this DC has enough capacity to accommodate the additional load. Similarly to the assignment of new load, periodically, when the difference between cost functions among the DCs grows above a given threshold, some workload can be migrated from the DC with the largest cost function to the other DCs. Different migration policies are possible with respect to the choice of VMs to be migrated.

3. PREDICTIONS

For the prediction of RE production, the neural network approach was taken into account. Feed Forward Neural Network (FNN) and Recurrent Neural Network (RNN) were compared, and after a preliminary test of different approaches, whose results are not reported here for the sake of brevity, RNN resulted to perform better than FNN in predicting RE production, thanks to its characteristics. Indeed, the main feature of RNNs is the presence of loops and their capability to store information about previous computations, which perfectly fits with our use case. In particular, by using Gated Recurrent Units (GRUs), the error in predicting RE production was lower than FNN; these networks have been used in modeling and prediction of sequential data in several applications that include image processing, sentiment analysis, language translation, and speech recognition.

The proposed RNN has one input layer with 10 neurons and one hidden layer with 11 neurons; these values were chosen to trade-off between prediction accuracy and complexity. Indeed, through the computation of the mean square error of the predictions, we could observe that the error is quite large when the number of neurons is smaller than 11, but remains almost the same when more than 11 neurons are used. Hence, 11 seems the smallest value that guarantees good accuracy.

The RNN predicts the RE production over time periods of one hour. The dataset used for training the RNN is the Photovoltaic Geographical Information System (PVGIS) [10, 11], an online free solar photovoltaic energy database. Most of the values of solar irradiance contained in the database derive from satellite data, which are used to estimate the solar radiation arriving at the earth surface; this procedure results to be more accurate than using sensors ground measurements. The method to compute solar radiation from satellite data is described in [12, 13]. Once the irradiance G has been predicted, the associated generated power, denoted P_G, is given by

$$P_G = G \sin(\theta) A \eta \qquad (2)$$

where θ is the PV panel inclination, set to 30°, A is the area of the panel and η is the panel efficiency. For training, we have used the data of the years from 2005 to 2015 and the samples of 2016 were used for testing. As an example to observe the accuracy of the predictions, Figure 1 reports the actual hourly production in blue and the predicted one in red during a week of January of the considered testing period. The peak production is not accurately predicted, however, the general shape is well caught.

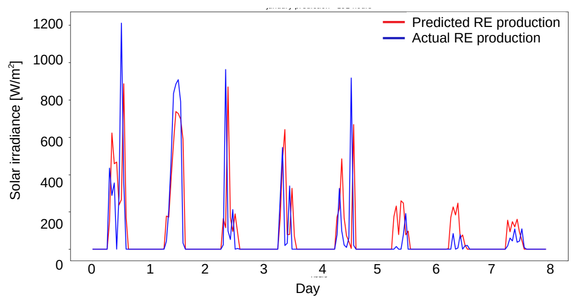


Fig. 1: Example of predictions

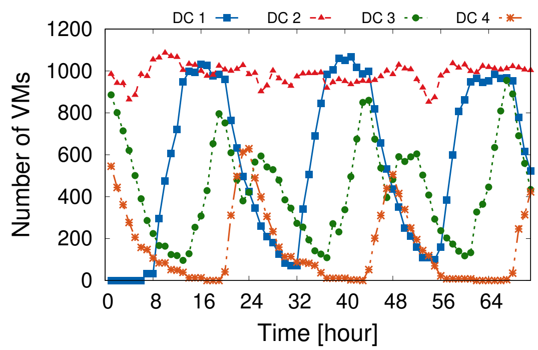


Fig. 2: Load per DC in Summer, assuming local RE production with PV panel size = 200 kWp and no storage.

4. RESULTS

In this section, we valuate the impact of predictions of RE production on the performance of the DC which runs the workload assignment strategy described above.

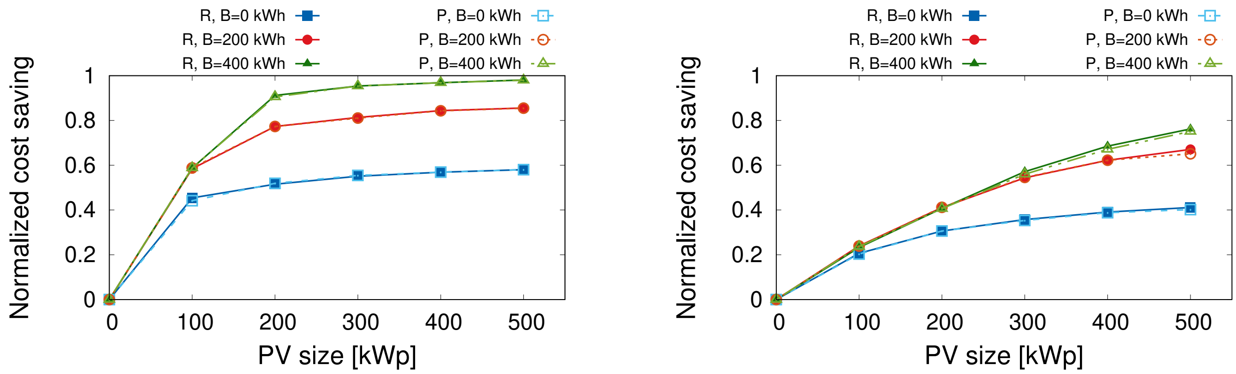
Figure 2 shows the load distribution among the DCs versus time for a few consecutive days. Different colors refer to different DCs. Observe how EcoMultiCloud adapts the load distribution to the conditions in the various DCs: the load reflects the daily RE production and the patterns are shifted according to the timezones. DC 4 is the least loaded, due to the high value of PUE.

Figure 3 reports the relative cost saving versus the size of the PV panels in summer and winter. Different values of the capacity of the batteries are considered. For each scenario, a dashed line reports the saving that is achieved when the RE production is predicted with the RNN described above; while the solid line is the case in which the workload management decisions are taken assuming an exact knowledge of the amount of RE that will be generated in the next hour. First, observe that, clearly, saving depends on the size of the PV panels. However, the saving does not significantly increase beyond a given size of the panels and the actual value of the saving depends on the battery size, indicating that the dimensioning of PV panels and batteries should be jointly defined to achieve the largest benefits. Second, notice that the cases of exact and predicted knowledge of the RE production basically coincide. The reason is that while predictions are needed to distribute the load based on a rough distinction among timezones and weather conditions, accurate values of the predictions are not strictly needed. This confirms that the proposed approach is feasible and effective also in the realistic assumption that a-priority knowledge of RE production is not possible. Finally, the large differences obtained in different seasons indicate that the evaluation of the benefit of introducing RE for powering big infrastructures requires to take into account seasonal variations.

Figures 4 and 5 compare the load distribution and cost function in two cases, the first without any migration policy (a) and the latter under the application of the random migration policy (b), meaning that the VMs to be migrated from a DC are randomly selected among the servers belonging to that DC and they are migrated towards the available DC having the lowest value of f_{assign}. Some differences can be observed in the number of VMs on the same DCs during the same period in the two tested cases. For example, between hours 8 and 12, DC 2 hosts more VMs in case (a) than in case (b), whereas in the latter case the number of VMs hosted by DC 3 and DC4 tend to be higher than in case (a). Indeed, as reported in Figure 5, the cost function f_{assign} for DC 3 and DC 4 has a lower value in case (b) with respect to case (a). In Figure 4b the dashed lines represent the DC load before the migration is performed, whereas the continuous lines correspond to the number of VMs per DC after the migration process has been completed. The curves are basically overlapping, indeed, the migration process is a frequent process that smoothly adjusts the load and speeds up the adaptation of load distribution to DC conditions. With the considered PV panel and storage sizing, under no migration policy, i.e., case (a), 23.9% of cost can be saved with respect to the case in which no local RE is produced, whereas cost can be reduced by 32.5% by applying the random migration policy as in case (b).

5. CONCLUSION

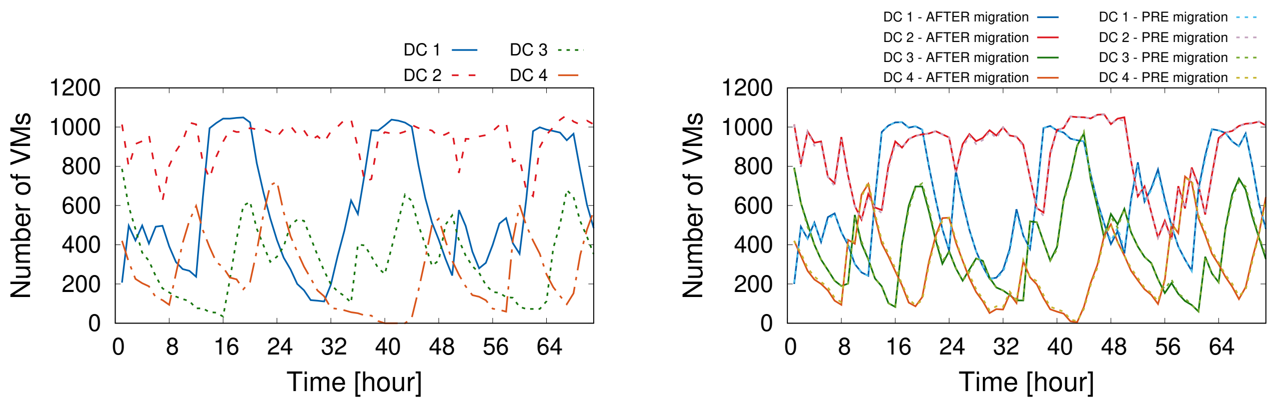
In this paper, we have considered geographically distributed DCs powered with hybrid systems that include, besides the connection to the power grid, RES and, in particular, PV panels. To properly operate, the workload management strategy needs predictions of the amount of RE that is generated by the PV panels. In the paper, we have shown that RNN can well predict the RE production. Interestingly, prediction errors at the peak values of the production are not critical because they do not compromise the load distribution decisions. Conversely, predictions are fundamental to distinguish, in an automatic way, the different production levels achieved in the various sites, levels that are influenced by the timezones and the weather conditions.



(a) Summer

(b) Winter

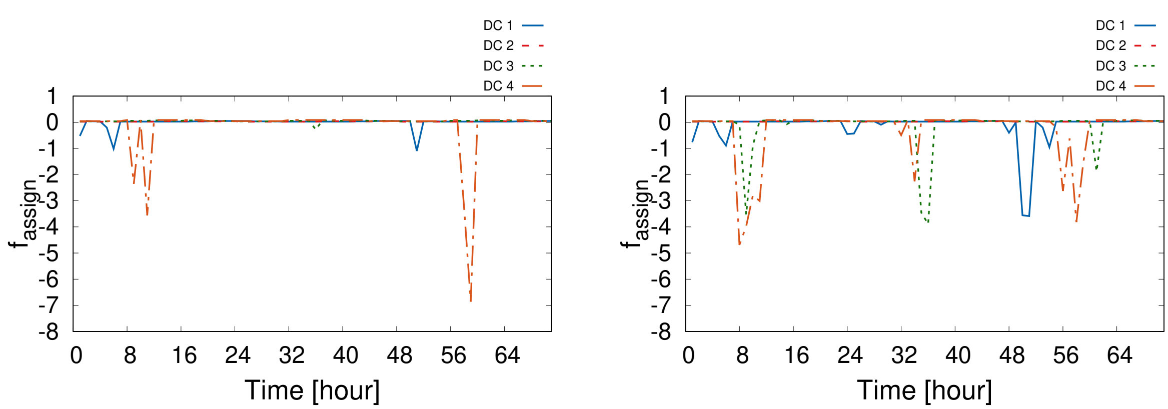
Fig. 3: Cost saving in summer and winter, assuming real (R) and predicted (P) RE production values, for different PV panel and battery sizes.



(a) No migration

(b) Random migration policy

Fig. 4: Load distribution per DC in Winter, assuming local RE production with PV panel size = 100 kWp and storage = 200 kWh, under no migration policy and under random migration policy.



(a) No migration

(b) Random migration policy

Fig. 5: f_{assign} per DC in winter, 100 kWp PV panel and 200 kWh storage, under no migration and random migration policy.

6. REFERENCES

[1] Y. Ran, H. Hu, X. Zhou, and Y. Wen, "Deepee: Joint optimization of job scheduling and cooling control for data center energy efficiency using deep reinforcement learning," in *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*, July 2019, pp. 645–655.

[2] A. Khosravi, L. L. H. Andrew, and R. Buyya, "Dynamic VM Placement Method for Minimizing Energy and Carbon Cost in Geographically Distributed Cloud Data Centers," *IEEE Transactions on Sustainable Computing*, vol. 2, no. 2, pp. 183–196, April 2017.

[3] D. Laganà, C. Mastroianni, M. Meo, and D.Renga, "Reducing the operational cost of cloud data centers through renewable energy," *MDPI Algorithms*, vol. 11, no. 10, 2018.

[4] Z. Liu, M. Lin, A. Wierman, S. Low, and L.L.H. Andrew, "Greening geographical load balancing," *IEEE/ACM Transactions on Networking*, vol. PP, no. 99, pp. 1–1, 2014.

[5] Asfandyar Qureshi, Rick Weber, Hari Balakrishnan, John Guttag, and Bruce Maggs, "Cutting the electric bill for internet-scale systems," *SIGCOMM Comput. Commun. Rev.*, vol. 39, no. 4, pp. 123–134, Aug. 2009.

[6] Dan Xu and Xin Liu, "Geographic trough filling for internet datacenters," in *2012 Proceedings IEEE INFOCOM*, Mar. 2012, pp. 2881–2885.

[7] Yuan Yao, Longbo Huang, Abhishek B. Sharma, Leana Golubchik, and Michael J. Neely, "Power cost reduction in distributed data centers: A two-time-scale approach for delay tolerant workloads," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 1, pp. 200–211, 2014.

[8] Yuanxiong Guo, Zongrui Ding, Yuguang Fang, and Dapeng Wu, "Cutting down electricity cost in internet data centers by using energy storage," in *2011 IEEE Global Telecommunications Conference (GLOBECOM 2011)*, Dec. 2011, pp. 1–5.

[9] A. Forestiero, C. Mastroianni, M. Meo, G. Papuzzo, and M. Sheikhalishahi, "Hierarchical approach for efficient workload management in geo-distributed data centers," *IEEE Transactions on Green Communications and Networking*, vol. 1, no. 1, pp. 97–111, March 2017.

[10] "Photovoltaic Geographical Information System. Overview of PVGIS data sources and calculation methods," https://re.jrc.ec.europa.eu/pvg_download/data_download.html.

[11] Thomas Huld, Richard Müller, and Attilio Gambardella, "A new solar radiation database for estimating PV performance in Europe and Africa," *Solar Energy*, vol. 86, pp. 1803–1815, 06 2012.

[12] Richard Müller, Christos Matsoukas, A Gratzki, Hein Behr, and R. Hollmann, "The CM-SAF operational scheme for the satellite based retrieval of solar surface irradiance - A LUT based eigenvector hybrid approach," *Remote Sensing of Environment*, vol. 113, pp. 1012–1024, 05 2009.

[13] Richard Mueller, Tanja Behrendt, Annette Hammer, and Axel Kemper, "A New Algorithm for the Satellite-Based Retrieval of Solar Surface Irradiance in Spectral Bands," *Remote Sensing*, vol. 4, no. 3, pp. 622–647, Mar 2012.