Greening Data Centers

Sara Alouf (Maestro, Inria) and Fabien Hermenier (Comred Project – Scale team, I3S)

SUMMARY

The objective of this thesis is to develop tools that will help greening data centers. The idea is to couple stochastic models for the harvested power and for the data center workload and use those to dimension the photovoltaic plant needed to achieve a predetermined goal. The contributions of the thesis will be both theoretical and experimental, aiming to provide insights into the following questions: how are current workloads and schedulers compatible with the intermittent nature of a photovoltaic power sources? What are the theoretical bounds that prevent to power an existing datacenter using photovoltaic arrays, coupled or not with power storage systems? How to size a photovoltaic array on top of an existing data center depending on the targeted environmental and economic objectives?

CONTEXT

Data centers are facilities that house a potentially huge amount of computer systems that are used mainly for storage or computing. The energy consumption of data centers is rapidly increasing making them the fastest-growing consumers of electricity in developed countries. Even though they are the largest in scale, cloud computing data centers are not the biggest contributors to the energy consumption. Most of the energy is consumed by smaller scale data centers. Therefore, we focus in this study on the latter category of data centers.

Typically, there are three subsystems in a data center: the equipment, the power infrastructure and the cooling infrastructure. To reduce the ecological footprint of data centers, one needs to optimize the three subsystems [4]. Better equipment may consume less power; using renewable energy reduces the electricity bill and improves sustainability; innovative solutions can improve the efficiency of the cooling infrastructure and reduce its power consumption. When the power infrastructure includes sources of renewable energy, it is highly interesting to shift as much of the load to the periods of high-energy production. This however may conflict with the best operation time seen by the cooling infrastructure. For instance, solar energy is abundant during clear days but at the same time, the temperature is at its highest point, which is the least convenient time to run servers at maximum load (the cooling infrastructure would need to be used at its maximum potential).

Sustainability in datacenters is currently addressed through energy-adaptive schedulers [2, 3, 4] or applications [5] that consider the energy as a resource similar to memory or CPU. This exhibits interesting results but tends to neglect the affinity that must exist between the workload and an intermittent powering source. Basically, some workloads naturally offer temporal and spatial manageability while others are too rigid to benefit from a daylight only energy production (like the photovoltaic energy).

Sizing the power source is also crucial to maximize its benefits. For example, an experiment within the DC4Cities project [5] runs on a cluster powered by either the grid or an ad-hoc local photovoltaic (PV) array. It appeared the array is two times too powerful with regards to the infrastructure needs. As a result, half the photovoltaic panels are turned off as the exceeding production cannot be stored or re-injected. Under-sizing the array reduces also its benefits as the switch that connects it to the electric supply only occurs when the array production covers the infrastructure consumption.

We think it is then meaningful to consider capacity-planning techniques to propose a sustainable powering architecture aligned with the data center characteristics and performance goals. For example, the usage of batteries for energy storage may be crucial in some regions to host particular workloads. Similarly, the number of photovoltaic panels should be evaluated carefully to estimate the most efficient architecture according to (i) the foreseen workload, (ii) the possible penalties when jobs are delayed, and (iii) the economic environment.

Thesis' objectives

The focus will be on theoretical and practical solutions to study the sustainability of existing data centers. A first objective is then to evaluate the scheduler and the workload of existing data centers to estimate their affinity with an intermittent power source. This evaluation will be performed both analytically and empirically. A second objective is then to rely on these results to assist data center owners at sizing renewable power sources through simulation.

For the analytical evaluation, stochastic models for the harvested power and for the data center workload will be coupled and used to dimension the photovoltaic plant needed to achieve a predetermined goal. The goal is to apply the models to data centers. These have both short and long jobs with potentially short deadlines that are scheduled on the fly. Therefore it is crucial to develop models on a small time-scale basis, typically minute-based for the harvested power. To this end, per-minute measurements of the solar irradiance (direct or diffuse) should be used. Such measurements are provided by NREL [6]. The thesis work can build upon the preliminary results of [7].

On the empirical side, traces of the jobs and of the energy consumption of up to three data centers (Grid'5000, Nef, and the cluster of University of Nice Sophia Antipolis) will be used to validate the developed models.

Challenges

Modeling the harvested power is very challenging. It is known that the solar irradiance exhibits a night-day behavior but it is also affected by weather conditions. Such conditions may induce burstiness in the energy produced. Models for the clear sky solar irradiance at zenith have been developed (see for instance [8]), but it remains a challenge to account for the actual irradiance received by PV panels at any time of the day. Recent studies have proposed analytical models for shaping the solar power; see [9, 10]. The models however rely on per-hour measurements which are not compatible with data centers workload characteristics.

A second challenge consists to propose a data center simulator that integrates precise energy provisioning and distribution models when connected to multiple sources (e.g. a grid and a local PV array). Despite the fact that data center simulators exist (see Simgrid [11]), their power models are related to the server and network consumption, and not the provisioning and the distribution aspects.

Thesis' outcome

The thesis will produce (1) stochastic models for the harvested power and for the jobs at the data center, (2) simulation tools to forecast the impact of PV arrays on top of existing data centers, and (3) a capacity planner to assist in the dimensioning of the PV array. Ultimately, the contributions of the thesis will provide insights into the following questions:

- How are current workloads and schedulers compatible with the intermittent nature of a photovoltaic power sources? What are the theoretical bounds that prevent to power an existing datacenter using photovoltaic arrays, coupled or not with power storage systems?
- How to size a photovoltaic array on top of an existing data center depending on the targeted environmental and economic objectives?

Advisors

Sara Alouf and Fabien Hermenier will share equally the supervision task. Currently none of them holds an HDR, but Sara is in the process of writing her HDR thesis and plans to defend it in the course of the year 2016.
Fabien is an expert in virtualization and resource management and has been working on virtual machine scheduling, green computing, and capacity planning [12]. Sara's skills in stochastic modeling and performance evaluation (e.g. [13]) will complement Fabien's expertise in this project. Their backgrounds cover the research topics that will be addressed in this thesis.

INTER-DISCIPLINE AND FIT IN THE LABEX UCN@SOPHIA

This thesis proposal combines the research topics of both the Scale and the Maestro teams. One of the research topics of Scale is resource management and sustainable computing. On the other side, Maestro is interested in the modeling and the performance analysis of networks and computer systems. The use of renewable energy in the IT sector has become a hot topic in the last years. Evaluating the sustainability of wireless networks has been a research topic for Maestro for a few years now, but studying specifically data centers powered by renewable energy sources is a new topic for Maestro (but not for our community as illustrated by the reference [4]). Within this thesis, the complementary expertise of both teams will be combined to achieve energy efficiency, a topic already addressed by the two teams but with different perspectives.

This thesis proposal is relevant to two of the research themes of the Labex UCN@Sophia, namely “distributed and ubiquitous computing” and “energy efficiency”.

POTENTIAL CANDIDATE

The two supervisors give lectures in the master 2 IFI Ubinet. One is related to performance evaluation; the other is related to resource management. The objective is to use these lectures to cast good students and confirm this casting by an internship. Traditional communication channels (CNRS mailing lists, public announce …) will also be used to attract other potential good candidates.

SOME KEY REFERENCES (NON-EXHAUSTIVE LIST)

[1] BtrPlace, An Open-Source flexible virtual machine scheduler, http://www.btrplace.org/.

[2] I. Goiri, W. Katsak, K. Le, T.D. Nguyen, R. Bianchini, “*Parasol and GreenSwitch: managing datacenters powered by renewable energy*,” Proc. of ASPLOS 2013, Pages 51-64.

[3] I. Goiri, K. Le, M.E. Haque, R. Beauchea, T.D. Nguyen, J. Guitart, J. Torres, R. Bianchini, “*GreenSlot: Scheduling energy consumption in green datacenters*,” Proc. of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis (SC), Seattle, Washington, USA, 12-18 Nov. 2011, pp.1-11.

[4] Z. Liu, Y. Chen, C. Bash, A. Wierman, D. Gmach, Z. Wang, M. Marwah, C. Hyser, “*Renewable and cooling aware workload management for sustainable data centers*,” Proc. of ACM SIGMETRICS/PERFORMANCE 2012. (ACM SIGMETRICS Performance Evaluation Review, Volume 40 Issue 1, June 2012, Pages 175-186).

[5] The EU DC4Cities Project: Environmentally sustainable data centre for smart cities. http://www.dc4cities.eu

[6] National Renewable Energy Laboratory (NREL). http://www.nrel.gov/.

[7] D. Politaki. “Stochastic Models for Solar Energy.” Master Thesis, University of Nice Sophia Antipolis, August 2014.

[8] R. E. Bird and R. Hulstrom. “*A Simplified Clear Sky Model for Direct and Diffuse Insolation on Horizontal Surfaces*.” Solar Energy Research Institute, Technical report SERI/TR-642-761, February 1981.

[9] M. Miozzo, D. Zordan, P. Dini and M. Rossi. “*SolarStat: Modeling Photovoltaic Sources through Stochastic Markov Processes*.” In Proc. of IEEE EnergyCon 2014.

[10] Y. Ghiassi-Farrokhfal, S. Keshav, C. Rosenberg, and F. Ciucu. “*Solar Power Shaping: An Analytical Approach*.” in IEEE Trans. on Sustainable Energy, Vol. 6, Issue 1, January 2015.

[11] H. Casanova, A. Giersch, A. Legrand, M. Quinson, and F. Sutero. “Versatile, Scalable, and Accurate Simulation of Distributed Applications and Platforms.” Journal of Parallel and Distributed Computing, 74(10):2899-2917, June 2014.

[12] F. Hermenier, R. Ricci. *How to build a better testbed: Lessons from a decade of network experiments on Emulab.* TridentCom 2012. **Best paper award.**

[13] I. Dimitriou, S. Alouf, A. Jean-Marie, “*A Markovian Queueing System for Modeling a Smart Green Base Station*.” In Proc. of EPEW 2015, Madrid, Espagne, August 2015.