2011 First International Conference on Data Compression, Communications and Processing

Saving Energy in Data Center Infrastructures

Sergio Ricciardi
Davide Careglio
Germán Santos-Boada
Josep Solé-Pareta
Dept. d'Arquitectura de Computadors
Universitat Politècnica de Catalunya
Barcelona, Spain
e-mail: sergior@ac.upc.edu

Ugo Fiore
Centro Servizi Informativi
Università di Napoli Federico II
Naples, Italy
e-mail: ufiore@unina.it

Francesco Palmieri
Dipartimento di Ingegneria dell'Informazione
Seconda Università di Napoli
Aversa, Italy
e-mail: fpalmier@unina.it

***Abstract*— At present, data centers consume a considerable percentage of the worldwide produced electrical energy, equivalent to the electrical production of 26 nuclear power plants, and such energy demand is growing at fast pace due to the ever increasing data volumes to be processed, stored and accessed every day in the modern grid and cloud infrastructures. Such energy consumption growth scenario is clearly not sustainable and it is necessary to limit the data center power budget by controlling the absorbed energy while keeping the desired level of service. In this paper, we describe EnergyFarm, a data center energy manager that exploits load fluctuations to save as much energy as possible while satisfying quality of service requirements. EnergyFarm achieves energy savings by aggregating traffic during low load periods and temporary turning off a subset of computing resources. EnergyFarm respects the logical and physical dependencies of the interconnected devices in the data center and performs automatic shut down even in emergency cases such as temperature peaks and power leakages. Results show that high resource utilization efficiency is possible in data center infrastructures and that huge savings in terms of energy (MWh), emissions (tons of CO_2) and costs (k€) are achievable.**

Energy-efficiency, power management, sleep mode, green data centers, grid computing, cloud computing.

I. Introduction

It is estimated that worldwide data centers alone consume 26 GW of electrical power corresponding to about 1.4% of the worldwide electrical energy consumption, with a growth rate of 12% per year [1][2]. To give an idea, the Barcelona Supercomputing Center (a medium-size data center) pays every year more than € 1 million just for the energy bill and consumes 1.2 MW [3], as much power as a town of 1,200 houses [4]. The power consumption in data centers originates from the involved computing, storage and interconnection equipment, together with the associated HVAC (heating, ventilation and air conditioning), UPS (uninterruptible power supply) systems and lighting facilities, with the servers being the most energy-hungry devices. The power usage effectiveness (PUE) index, defined by the Green Grid [5], measures the efficiency of an ICT facility as the ratio of the total amount of power used by the facility to the power delivered to the computing equipment alone. While larger data centers tend to be able to implement more efficient cooling, high availability needs may require the use of expensive UPS and more redundancy, which then result in a higher PUE. A PUE value of 2 is the current average [6], meaning that HVAC and UPS double the energy requirements. In data centers, a wide variety of computing resources are usually available, ranging from small servers with computational capabilities comparable to personal computers, to large supercomputers. Furthermore, there are different types of servers optimized for specific tasks such as web and database servers. One of the largest problems of this equipment is the relative independence of the power consumption with their real operating load [6] and the consequent need for energy-proportional architectures [7]. This, combined with the fact that many servers are being operated far below their actual capacity [7][8], leads to a lot of wasted energy in data centers, thus enabling great potential energy savings. Next to using optimized components, a second level of optimization lies in power management. In such a scenario, three energy-saving approaches are available: "do less work", "slow down" and "turn off idle elements". In the "do less work" strategy, the processes are optimized so that the load to be executed becomes minimal, resulting in lower power consumption. The "slow down" strategy considers that the faster a process runs, the more resource intensive it becomes. In complex processes, the speeds of several sub-processes don't match and thus resources are used without being absolutely required. There are two ways of slowing down processes. They can be run with adaptive speeds, by selecting the minimal required speed to complete the process in time. Alternatively, buffering can be introduced so that instead of running a process immediately upon arrival, one can collect new tasks until the buffer is full and then execute them in bulk. This allows for components to be temporarily switched off resulting in lower power consumption. The "turn off idle elements" strategy refers to the possibilities offered by exploiting a low-consumption state (sleep mode). Basically, the sleep mode aims at switching into an idle mode the devices during periods of inactivity. Unloaded servers can be dynamically put into sleep mode during low-load periods, contributing to great power savings. For data centers and grid/cloud infrastructures, if properly employed, the sleep mode may represent a very useful mean for limiting power

978-0-7695-4528-8/11 $26.00 © 2011 IEEE
DOI 10.1109/CCP.2011.9

IEEE computer society

consumption of lightly loaded sites. Data centers are in fact inherently modular, as they are built up by a number of logical-equivalent elements (bulks of servers). A grid site, for example, is basically composed by a disk pool manager (DPM) that controls data storage (storage elements SE, disk servers DS, storage systems SS), and a computing element (CE) that sends jobs to working nodes (WN). While the DPM and the CE are usually hosted on individual servers (for a grid of medium size), the SEs and especially the WNs functions may be distributed over a very large number of nodes. All these servers are up and running even if the farm is scarcely loaded or idle. Aggregating the jobs on a subset of SEs and WNs allows putting into sleep mode all the remaining nodes greatly reducing the energy consumption (assuming replicated data on SEs). In this direction, we started from the PowerFarm software [9] developed to manage power losses and temperature/humidity peaks in grid sites, that can be used to automatically shut down devices in case of emergency (such as temperature peaks, smoke or fire alerts, etc). It works by processing the SNMP trap alerts and taking the corresponding preconfigured actions, but it lacks the intelligence to take any energy saving action. Thus, we extended the PowerFarm framework to monitor current loads and server power consumptions and to turn on/off servers as needed while respecting physical and logical dependencies among them. Accordingly, we developed EnergyFarm, a simple and effective energy control system, which, through a service-demand matching algorithm, determines the subset of servers that may be powered off while satisfying the data center computing and storage demand.

II. RELATED WORK

In order to reduce the energy consumption of data centers, a number of directions have been highlighted in the literature. In [1] it is argued that significant power savings can be realized through virtual server configurations, allowing to switch off most servers during night hours and only using the full capacity of servers during peak hours. In [10], a number of measures are identified: legacy equipment requiring appropriate software may undergo hardware upgrades (such as modified power supply modules) and their network presence may be transferred to a proxy or agents allowing the end device to be put in low consuming mode during inactivity periods while being virtually connected to the Internet. The authors also plead the need to enable renewable energy sources, such as solar, wind or hydro power, to supply power to ICT systems. This approach seems specifically applicable to data centers, which can be located at renewable energy production sites. However, since renewable energy sources tend to be unpredictable (e.g., wind), or vary during day and night (e.g., sun), this would imply that the data itself need to be migrated from one data center to the other, according a so-called follow-the-sun or chase-the-wind scenario [11]. As a consequence, energy-efficient high bandwidth networks and routing architectures will be required. Along this line of thought, a study performed in [12] investigates cost-aware and energy-aware load distribution across multiple data centers. The study evaluates the potential cost and carbon savings for data centers located in different time zones and partly powered by green energy and founds that, when optimizing for green energy use, green data centers can decrease CO_2 emission by 35% by leveraging the green data centers at only a 3% cost increase. Several sources [1][6][13][14] in literature have pointed out the sleep mode as a solution for achieving energy-efficiency. In [6] the authors focus on component level where more efficient technologies should be used. In [15] a network power manager is presented, which dynamically adjusts the set of active network elements (links and switches) to satisfy changing data center traffic loads; it is focused on the network infrastructure of the data centers. Our work is instead focused on improving the operating energy efficiency of the computing resources (servers), which are responsible for the greatest part of data centers energy consumption.

III. AN ENERGY-AWARE DATA CENTER CONTROL PLANE

A data center is composed by a number of servers running jobs (or tasks) that come from the Internet. Every server has a processing capacity, depending essentially on the number of cores and/or processors. The data center *workload* is thus represented by the jobs that the data center has to process in each moment. Typical data centers are strongly over-provisioned to work well under peak workloads [7][8]. However, idle servers are normally kept turned on even if there are no jobs to process. This clearly represents a waste in the power utilization and a cost in the energy bill. Our goal is to reduce the set of active servers to a subset of servers and turn off the idle ones, according to the "turn off idle elements" approach. In this scenario, EnergyFarm will be the high-level energy-aware control plane logic that complements and extends the low-level PowerFarm actuator facilities, which physically manage the power distribution in the data center. In order to exploit load fluctuations by turning off inactive servers and saving energy, we defined a specific operating *policy* within EnergyFarm that establishes *what* should be done, and implemented the corresponding *mechanisms* in PowerFarm which specify *how* it should be realized. Thus, as the job traffic load changes, the policy indicates which servers have to be turned off and the PowerFarm facilities implement the correct procedures to accomplish the task while respecting the physical and logical dependencies among the data center devices. In particular, the main EnergyFarm operating policy has been implemented through a proactive algorithm – running within the farm resource broker – that constantly monitors the traffic load and dynamically decides the subset of servers that may be turned on/off, while the PowerFarm actuator functions have the task to correctly power on/off such servers.

A. *Modeling Resource Allocation and Traffic Fluctuations*

In our model, we follow the usual scenario in which each job is assigned to one CPU core, so that a server with multiple CPUs making available n total cores may run n jobs without experiencing any performance slowdown. For multi-core CPUs, we take advantage of such characteristic by aggregating jobs on a subset of servers in order have more idle servers to turn off. For servers with n CPU cores, several aggregation strategies are possible: among the active servers, first-fit assigns a new job to the first server with one CPU core available. Best-fit tries to compact the jobs as much as possible; the new job is assigned to a server with just 1 core free (and, thus, $n - 1$ busy) if any such server exists. Otherwise, it looks for a server with 2 free cores, then with 3, and so on, up to n. Clearly, first-fit is faster but it leaves a great number of servers not fully loaded; best-fit gives the best results since it compacts the jobs as much as possible and frees the maximum number of servers that may be turned off. Besides compacting as much as possible, best-fit is also profitable since a multi-core server with a high number of busy cores is less probable to get free of all his jobs (and, thus, of being put into sleep mode) than a server with a low number of jobs. Best-fit computational complexity may be improved to work in constant amortized time by implementing the server priority queue with a Fibonacci heap. If servers are single-core devices, no aggregation is possible and all the energy savings come exclusively from the shutdown of the idle servers.

Typical data centers traffic demands are not constant over time: on the contrary, they are characterized by high utilization periods (e.g., during some hours of the day) followed by low utilization periods (e.g., during the night). In particular, it has been observed that the traffic load fluctuations are almost predictable within certain fixed time periods (e.g., day-night, months or years) and resemble a pseudo-sinusoidal trend [15][16]. In Fig. 1 it is reported the theoretical daily traffic variation of a typical energy-unaware production site [15]: the traffic load (demand) follows a pseudo-sinusoidal trend whilst the power keeps constant during high and low load periods. This behavior is due to data center resources that are always on and consume energy even during low load periods. The idea is to introduce *elasticity* in the demand-capacity provisioning, by dynamically varying the capacity with the demand, like depicted in Fig. 2. In theory, the capacity should resemble the demand as closely as possible, but two main problems have to be addressed. First, the capacity is not a continuous curve but is instead a step function in which each step corresponds to a computing resource (e.g., a server in the farm) turned on/off. Thus, the demand curve has to be approximated with a step service curve that serves the demand while minimizing the energy consumption (Fig. 3). Second, the provisioned capacity should have a *safety margin* (i.e., a distance d between the demand and the capacity curves) to cope with peak loads. The margin represents the number of servers that are preventively turned on for serving new jobs to come. The smaller the d, the lower the energy consumption, but also the lower the number of jobs that will be served without delay. The higher the d value, the more the jobs that will be served as they arrive, but also the higher the energy consumption (since a greater number of servers have to be powered on). The safety margin d has to be large enough to avoid oscillating between states for little variations of the load. During the start up of a server, in fact, peaks in the power absorption are experienced, due to the server bootstrap procedure and the OS loading process. Therefore, d is upper-bounded by the energy consumption and lower-bounded by the peak load absorption capacity and oscillation minimization requirement. At any instant, the absorbed power is directly proportional to the number of active servers; so the closer the service curve resembles the demand curve, the lower the required power will be. With the safety margin d, a bulk of $k \leq d$ incoming jobs will not have to wait. Thus, the d parameter sets the size of the zero-waiting queue of jobs that are immediately served as they arrive. If $k > d$, there will be $k - d$ jobs that will have to wait a time t before they can get served, where t is the start-up time of the servers (obviously, if the load reaches the site maximum capacity, all new jobs will have to wait for new resources to become available). The start-up time t may sensibly vary with the available technology. For agile servers equipped with enhanced sleep mode capabilities, t may be in the order of *ms*, whilst for legacy equipment a complete bootstrap procedure will be required and t may grow up to some minutes (see Table I). In general, the higher the t value, the higher the d, and thus the lower the energy saving margin, while with low values of t, greater energy savings are possible. In Table I we reported the (software and hardware) turn off and wake-up times measured in the INFN[1] Tier2 Grid Site of the CERN[2] LHC[3] experiments. Legacy servers, not equipped with the sleep mode, need several tens or even hundreds of seconds to switch state. Such high times indicate that the enhanced sleep mode feature is strongly advised and may bring great benefits in terms of energy savings, as the results in Section IV confirms.

TABLE I. COMPLETE TURN ON/OFF TIMES (SECONDS) FOR DIFFERENT DEVICES.

Server type	*Power on (hardware)*	*Power off (software)*	*Power off (hardware)*
Computing Element (CE)	120	20	5
Storage Element (SE)	180	10	5
Home Location Register (HLR)	120	60	5
Pizzabox form factor Servers	120	10	5
Blade Servers (Dell® DRAC)	160	45	45
Storage Server (IBM® DS400 Storage System)	60	10	10

B. Energy Savings Potential

In order to evaluate the maximum potentialities of our energy saving approach, we consider instantaneous transitions among the sleep and the active states ($t = 0$) and theoretical sinusoidal traffic, like the one depicted in Fig. 3. The demand curve represents the traffic load during the day, while the service curve represents the servers that must be

[1] Italian National Institute for Nuclear Physics, Naples, Italy.

[2] European Organization for Nuclear Research, Geneve, France-Switzerland.

[3] Large Hadron Collider.

active to process the job requests. Without any energy saving management, the power consumption of the data center stays constant [15], and the energy is the integral of power over time:

$$\int_{t_1}^{t_2} p(t)dt \qquad (1)$$

where *p(t)* is the power consumption function and t_1 and t_2 are the considered time extremes. Ideally, the lower bound for the data center energy consumption is given by:

$$\int_{t_1}^{t_2} l(t)dt \qquad (2)$$

where the *l(t)* function describes the load curve. EnergyFarm approximates such curve with the service curve *s(t)*, which is the step function that establishes the minimum set of resources that have to stay powered on to serve the current demand. Therefore, with our energy saving schema the theoretical energy consumption is given by:

$$\int_{t_1}^{t_2} s(t)dt\ . \qquad (3)$$

Clearly, it holds that (2) < (3) << (1), and the bigger the difference between (1) and (3) the greater the energy saving. Theoretically, the energy saving is upper-bounded by:

$$\int_{t_1}^{t_2} (p(t) - l(t))dt \qquad (4)$$

while the actual energy saving is given by:

$$\sum_{i=1}^{n} (p(i) - s(i)) \cdot \Delta i \qquad (5)$$

where n is the number of intervals in which the time interval $[t_1, t_2]$ is divided and Δi is the duration of the i-th time interval; note that n sets the time-basis on which the EnergyFarm is executed. Therefore, eq. (5) represents the energy saving of our EnergyFarm approach which will be evaluated in detail in Section 6.

C. *The service-demand matching algorithm*

Given a demand curve, the EnergyFarm service-demand matching algorithm determines the service curve that satisfies the demand while limiting the number of active server and, thus, the power consumption. As an example, let's consider the scenario depicted in Fig. 4 and Fig. 5.

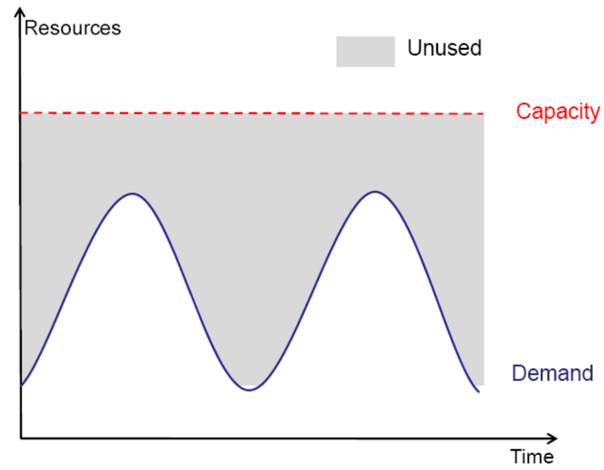

Figure 1. Capacity-demand mismatch leads to resource and energy wastes.

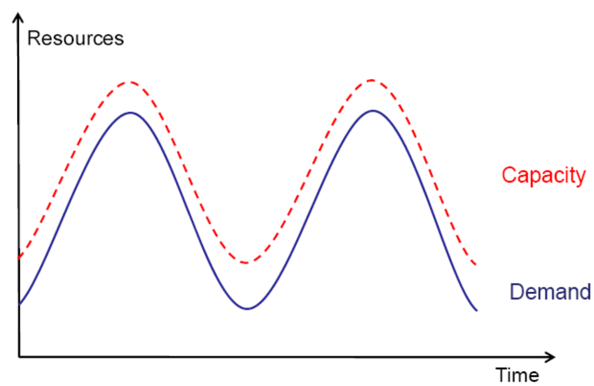

Figure 2. Theoretical provisioning elasticity concept.

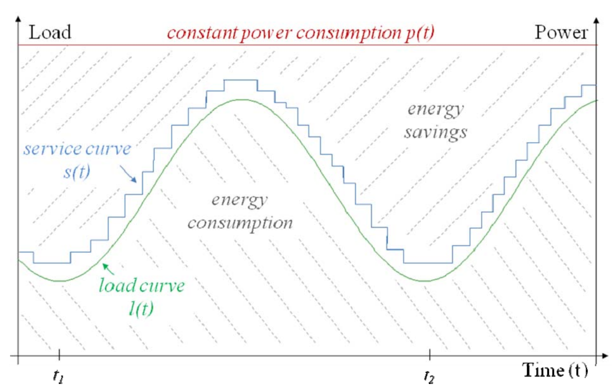

Figure 3. Service-demand matching.

In Fig. 4 the demand curve increases between t_i and t_{i+1}, consequently the distance from the service curve decreases from d_i to d_{i+1}. Since $d_{i+1} < d$, the algorithm detects the increase in the demand (totally absorbed by the guard band d, thus no delay is added in this case) and consequently increases the number of active servers by turning on $s_{i+1} - s_i$ servers.

Figure 4. Incrementing the service curve in response to demand increase.

The opposite situation is depicted in Fig. 5, where a decrement $d_{i+1} > d$ in the demand curve causes the algorithm to decrease the service curve from s_{i+1} to s_i.

Figure 5. Decrementing the service curve in response to demand decrease.

IV. PERFORMANCE ANALYSIS RESULTS

We evaluated the performances of EnergyFarm through simulations against several available data referring to two different-sized data centers. To model a large data center we used the Google farm [7][8] composed by more than 5,000 blade servers monitored over a six-month period; for the small data center we used the Naples LHC Tier2 Grid site of the INFN section [9] composed by more than 100 servers in the pizzabox form factor. First, we evaluated the impact of the safety margin *d* against the potential savings, in terms of energy (MWh), emissions (tons of CO_2) and economical cost (k€). For a commercial/industrial facility like a data center, the average cost of energy is € 0.12 per kWh [17]. We considered fossil-fueled energy plants powering the data centers, which emits 890 grams of CO_2 per kWh (ACV-DRD study [1]). Several simulations have been conducted for different values of the safety margin *d*. Results show that, for the small data center, the maximum cost savings is more than 35 k€ per year, while for the large data center the cost saving may reach € 1.5 millions, with a reduction of more than 13 GWh in the energy consumption and more than 11 kTons of CO_2 in the emissions (see Table II). Such results should not surprise: servers are rarely utilized at their full capacity and most of the time operate at between 10% and 50% of their maximum utilization levels [7]. As expected, the *d* value affects the energy savings and the consequent CO_2 emissions and bill costs. Best results have been achieved with low values of the safety margin.

TABLE II. PER YEAR SAVINGS WITH ENERGYFARM (SINGLE-CORE SERVERS) AND VARIABLE SAFETY MARGINS *d*.

Safety margin	***Energy (MWh)***	***Emissions (Tons of CO2)***	***Cost (k€)***
Small data center			
d = 1%	299.2	266.2	35.9
d = 10%	259.9	231.3	31.2
d = 50%	92.2	82.1	11.1
Large data center			
d = 1%	13184.9	11735.0	1582.2
d = 10%	11455.3	10195.3	1374.6
d = 50%	4065.3	3618.1	487.8

Note that, since our goal is to provide a lower bound for the energy savings of the modern and future data center, the transition time *t* between the on and off states have been put to 0, thus there is no delay in the powering on/off the servers (i.e. all agile servers). As a consequence, the frequency of the load variations (i.e., how and how often the traffic load varies in time) only affects the number of transitions between on/off states, but it does not influence the energy savings at all, as each variation is immediately followed by the corresponding on/off action on the involved servers. In our tests, the efficiency in the resource utilization has reached similar values for the small and the large data centers, varying from 20% to 68%, meaning that a good percentage of the servers has been put into sleep mode for considerable time (see Table III).

TABLE III. ENERGYFARM EFFICIENCY IN THE RESOURCE UTILIZATION.

Average resource utilization in EnergyFarm (%)		
d = 1%	***d = 10%***	***d = 50%***
68,0282	59,1045	20,975

The EnergyFarm saving margins decrease almost linearly as the *d* values increases (Fig. 12). In fact, while the load is far from the actual data center capacity, savings and *d* vary linearly but, as load approaches higher values, the threshold *d* will exclude a higher number of devices from being turned off, leading to relatively lower savings. When considering multi-core devices, job aggregation is possible. Two aggregation strategies have been studied: first-fit and best-fit. In our tests first-fit has always performed worse than best-fit (up to 50%), so here we only focus on the best-fit strategy. Results, both for the small and large data centers, show a common behavior, even if with different rates: the more the cores in the data center, the more the energy consumption. This is due to the fact that the data centers work far from their actual maximum utilization capacity and causes multi-core server to operate with only few jobs even with the best-fit strategy, i.e. multi-core servers present internal fragmentation (not all the cores are always busy).

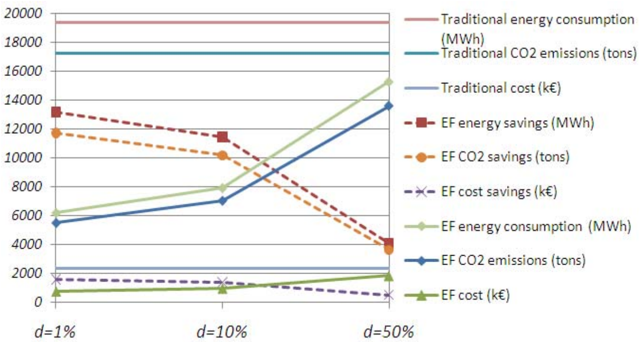

Figure 6. Energy, CO_2 and costs with varying d values (large data center).

Thus, at low loads, assigning one job to a single-core server costs less than executing it on a 8-core server (due to the greater energy consumption of the latter), whilst, at higher loads, the greater computing density of multi-core servers may be exploited by the best-fit strategy to lower the overall data center energy consumption.

TABLE IV. ENERGYFARM PERFORMANCES WITH VARYING NUMBER OF CORES PER SERVERS (d = 1%).

Cores per server	***1***	***2***	***4***	***8***
Aggregation	No	Best-fit		
Small Data Center				
Energy (MWh)	see Table II	138.36	142.79	152.05
CO2 (Tons)		123.14	127.08	135.32
Cost (k€)		16.60	17.13	18.25
Large Data Center				
Energy (MWh)	see Table II	6003.57	6007.36	6015.16
CO2 (Tons)		5343.18	5346.55	5353.49
Cost (k€)		720.43	720.88	721.82

V. CONCLUSIONS

In this work, we presented EnergyFarm, an energy manager which can be used on the modern and future grid/cloud data center infrastructures to save energy. Current farms are usually over-provisioned and fluctuations in the traffic load are observed at various time periods. To take advantage of such a situation, we developed EnergyFarm which, through the service-demand matching algorithm and the job aggregation capabilities, allows turning off idle servers, while respecting both the demand requirements and the logical and physical dependencies. Results showed that great efficiency in the resource allocation can be achieved (between 20% and 68%), allowing significant energy, cost and emissions savings. In the optic of the future ICT developments, EnergyFarm may become an indispensable instrument towards sustainable society growth and prosperity.

ACKNOWLEDGMENTS

This work was in part supported by the COST Action IC0804 on Energy Efficiency in Large Scale Distributed Systems, the Spanish Ministry of Science and Innovation under the DOMINO project (TEC2010-18522), the Catalan Government under the contract SGR 1140 and the DIUE/ESF under the grant FI-201000740.

REFERENCES

[1] BONE project, "WP 21 Topical Project Green Optical Networks: Report on year 1 and updated plan for activities", NoE , FP7-ICT-2007-1 216863 BONE project, Dec. 2009.

[2] J. Koomey, "Estimating Total Power Consumption by Servers in the U.S. and the World", February 2007, http://enterprise.amd.com/Downloads/svrpwrusecompletefinal.pdf.

[3] Jordi Torres, "Green Computing: the next wave in computing", Ed. UPCommons, Technical University of Catalonia (UPC), February 2010, Ref. http://hdl.handle.net/2099.3/33669.

[4] Peter Kogge, "The tops in flops", pp. 49-54, IEEE Spectrum, Feb. 2011.

[5] The Green Grid, "The Green Grid Data Center Power Efficiency Metrics: PUE and DCiE", Technical Committee White Paper, 2008.

[6] W. Vereecken, W. Van Heddeghem, D. Colle, M. Pickavet, P. Demeester, "Overall ICT footprint and green communication technologies", in Proc. of ISCCSP 2010, Limassol, Cyprus, Mar. 2010.

[7] L.A. Barroso, L. A., Hölzle, U., "The Case for Energy-Proportional Computing", IEEE Computer, vol. 40, 33-37, 2007.

[8] X. Fan, W.-D. Weber, and L.A. Barroso, "Power Provisioning for a Warehouse-Sized Computer", http://research.google.com/archive/power_provisioning.pdf.

[9] A. Doria, G. Carlino, S. Iengo, L. Merola, S. Ricciardi, M. C. Staffa, "PowerFarm: a power and emergency management thread-based software tool for the ATLAS Napoli Tier2", proceedings of Computing in High Energy Physics (CHEP) 21 - 27 March 2009, Prague, Czech Republic.

[10] W. Van Heddeghem, W. Vereecken, M. Pickavet, P. Demeester, "Energy in ICT - Trends and Research Directions", in Proc. IEEE ANTS 2009, New Delhi, India, Dec. 2010.

[11] B. St Arnaud, "ICT and Global Warming: Opportunities for Innovation and Economic Growth", http://docs.google.com/Doc?id=dgbgjrct_2767dxpbdvcf.

[12] K. Ley, R. Bianchiniy, M. Martonosiz, T. D. Nguyen, "Cost- and Energy-Aware Load Distribution Across Data Centers", SOSP Workshop on Power Aware Computing and Systems (HotPower '09), Big Sky Montana (USA), 2009.

[13] M. Gupta, S. Singh, "Greening of the internet", in Proc. of the ACM SIGCOMM, Karlsruhe, Germany, 2003.

[14] K. Christensen and B. Nordman, "Reducing the energy consumption of networked devices", in IEEE 802.3 tutorial, 2005.

[15] B. Heller, S. Seetharaman, P. Mahadevan, Y. Yiakoumis, P. Sharma, S. Banerjee, N. McKeown, "Elastictree: Saving energy in data center networks", in Proceedings of the 7th USENIX Symposium on Networked System Design and Implementation (NSDI), pages 249--264. ACM, 2010.

[16] M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, M. Zaharia, "Above the Clouds: A Berkeley View of Cloud computing", Technical Report No. UCB/EECS-2009-28, University of California at Berkley, USA, Feb. 10, 2009.

[17] U.S. Energy Information Administration, State Electricity Prices, 2006, http://www.eia.doe.gov/neic/rankings/stateelectricityprice.htm.