CEPIS UPGRADE is the European Journal for the Informatics Professional, published bi-monthly at <http://cepis.org/upgrade>

Publisher
CEPIS UPGRADE is published by CEPIS (Council of European Professional Informatics Societies, <http://www.cepis.org/>), in cooperation with the Spanish CEPIS society ATI (*Asociación de Técnicos de Informática,* <http://www.ati.es/>) and its journal Novática

CEPIS UPGRADE monographs are published jointly with Novática, that publishes them in Spanish (full version printed; summary, abstracts and some articles online)

CEPIS UPGRADE was created in October 2000 by CEPIS and was first published by Novática and INFORMATIK/INFORMATIQUE, bimonthly journal of SVI/FSI (Swiss Federation of Professional Informatics Societies)

CEPIS UPGRADE is the anchor point for UPENET (UPGRADE European NETwork), the network of CEPIS member societies' publications, that currently includes the following ones:
- inforewiew, magazine from the Serbian CEPIS society JISA
- Informatica, journal from the Slovenian CEPIS society SDI
- Informatik-Spektrum, journal published by Springer Verlag on behalf of the CEPIS societies GI, Germany, and SI, Switzerland
- ITNOW, magazine published by Oxford University Press on behalf of the British CEPIS society BCS
- Mondo Digitale, digital journal from the Italian CEPIS society AICA
- Novática, journal from the Spanish CEPIS society ATI
- OCG Journal, journal from the Austrian CEPIS society OCG
- Pliroforiki, journal from the Cyprus CEPIS society CCS
- Tölvumál, journal from the Icelandic CEPIS society ISIP

Cover page designed by Concha Arias-Pérez
"Luminous Recharge" / © ATI 2011
Layout Design: François Louis Nicolet
Composition: Jorge Llácer-Gil de Ramales

Editorial correspondence: Llorenç Pagés-Casas <pages@ati.es>
Advertising correspondence: <info@cepis.org>

Subscriptions
If you wish to subscribe to CEPIS UPGRADE please send an email to info@cepis.org with 'Subscribe to UPGRADE' as the subject of the email or follow the link 'Subscribe to UPGRADE' at <http://www.cepis.org/upgrade>

ISSN 1684-5285

Monograph of next issue (December 2011)
"**Risk Management"**

CEPIS UPGRADE

The European Journal for the Informatics Professional
http://cepis.org/upgrade

Vol. XII, issue No. 4, October 2011

Monograph

Green ICT: Trends and Challenges

(published jointly with Novática*)

Guest Editors: *Juan-Carlos López-López, Giovanna Sissa, and Lasse Natvig*

UPENET (UPGRADE European NETwork)

CEPIS NEWS

* This monograph will be also published in Spanish (full version printed; summary, abstracts, and some articles online) by **Novática**, journal of the Spanish CEPIS society ATI (*Asociación de Técnicos de Informática*) at <http://www.ati.es/novatica/>.

A State-of-the-Art on Energy Efficiency in Today's Datacentres: Researcher's Contributions and Practical Approaches

Marina Zapater-Sancho, Patricia Arroba-García, José-Manuel Moya-Fernández, and Zorana Bankovic

Energy efficiency has become an issue of great importance in today's datacentres. Metrics like Top500, which measure speed and performance, are beginning to lose importance in favor of others such as Green500. In order to increase energy efficiency of datacentres and save energy costs, the research community proposes solutions from both the computing and the cooling point of view, while European and US Institutions publish best practice manuals on energy-efficiency for datacentre owners. However, even though best practices are beginning to be implemented, most of the solutions offered by researchers are not yet used in real production environments. This paper makes a survey of the solutions proposed by researchers as well as the practices that real datacentres apply in order to increase the energy-efficiency of their facilities, and to find the reasons that create this gap between research and innovation in datacentres.

Keywords: Cooling, Datacentres, Energy Efficiency, Green Computing, Heterogeneous Systems, High-Performance Computing, Power Usage Efficiency (PUE), Reliability, Thermal-Aware Scheduling.

1 Introduction

In today's world, immersed in the Information Society, the driving force of economy has changed from being fossil fuels and electricity to being information. But the usage of Information and Communications Technologies has not only been restricted to the economy field; it has extended to all areas of society. Computers and electronic devices are found in almost all areas of human activity.

Within this framework, datacentres have turned into a key element in society. It is easy to find datacentres in all economic sectors, as they provide the computational infrastructure for a wide range of applications and services, covering from the specific needs of social networks to the requirements of high performance computing. Moreover, concerns over the reliability of these datacentres, understood as a concern about the availability and security of these infrastructures, from the point of view of the backup of information, is gaining greater importance.

The most valuable possession of companies is information; this means that datacentres are supposed to be always available, with all their servers up and running and all information appropriately secured. A loss of information or the lack of availability could result in high economic costs.

In this paper, we will study the state-of-the art on energy efficiency of today's datacentres. In Section 2 we will

Authors

Marina Zapater-Sancho received a Degree in Telecommunication Engineering and a Degree in Electronic Engineering from the *Universitat Politècnica de Catalunya*, Spain, in 2010. Currently she is a PhD Candidate in the Electronic Engineering Dept. at the *Universidad Politécnica de Madrid*, Spain. She has been awarded a Research Grant by the Program for Attracting Talent (PICATA) of the Campus of International Excellence of Moncloa. Her research focuses on thermal optimization of complex heterogeneous systems, proactive and reactive thermal-aware optimization of datacentres and security of embedded systems. <marina@die.upm.es>

Patricia Arroba-García received a Degree in Telecommunication Engineering from the *Universidad Politécnica de Madrid*, Spain, in 2011. Currently she is doing her PhD in the same University in the Electronic Engineering Dept. Her research field focuses on thermal optimization of complex heterogeneous systems and mobile accessibility development. <parroba@die.upm.es>

José-Manuel Moya-Fernández is currently an Associate Professor in the Dept. of Electronic Engineering, *Universidad Politécnica de Madrid*, Spain He received his MSc and PhD in Telecommunication Engineering from the same University, in 1999 and 2003, respectively. He has participated in a large number of national research projects and bilateral projects with industry, in the fields of embedded system design and optimization, security optimization of embedded systems and distributed embedded systems. In these fields he is co-author of a number of publications in prestigious journals and conferences. His current research interests focus on proactive and reactive thermal-aware optimization of datacentres, and design techniques and tools for energy-efficient compute-intensive embedded applications. <josem@die.upm.es>

Zorana Bankovic got the title of Electrical Engineer from the Faculty of Electrical Engineering, University of Belgrade, Serbia, in 2005, and the PhD title from the *Universidad Politécnica de Madrid*, Spain, in 2011. Currently she is a researcher in the Dept. Electronic Engineering at the *Universidad Politécnica de Madrid*. Her main research interests include reputation systems and energy efficient intrusion detection for wireless sensor networks based on artificial intelligence techniques. <zorana@die.upm.es>

> “This paper makes a survey of the solutions proposed by researchers as well as the practices that real datacentres apply in order to increase the energy-efficiency of their facilities”

present the main problems, then we will study the present solutions from a dual perspective. Section 3 reviews the actual trends for energy efficiency in high-performance computing datacentres, whereas Section 4 expands that study towards datacentres for cloud computing. Both these sections will make a survey from the research point of view. Section 5 will study how many of the technological advances offered by research in this area have a real market penetration, and which are the real solutions implemented in today's datacentres. Finally, some conclusions about the actual gap between research and industry in this area are drawn in Section 6.

2 Problem Statement

2.1 The Energy Problem

For decades, datacentres have focused on increasing their performance, defined only in terms of speed. Examples include the Top500 list of the world's fastest supercomputers [1], which calculates the speed metric as floating-point operations per second (flops). However, this speed increase has not come for free, as the energy efficiency has not improved as quickly. In 2007, although there had been a 10,000-fold increase in speed since 1992, performance per watt had only improved 300-fold and performance per square foot only 65-fold [2]. The significant growth in processing capabilities has lead to a severe rise in the energy consumption by these infrastructures, not only to increase computational power but also for to run more powerful cooling equipment. The energy consumption of datacentres has a global impact and in 2009 used the 2% of the world electric energy production [3].

The huge performance improvement is mainly due to increases in three different dimensions: the number of transistors per processor, each processor's operating frequency, and the number of processors in the system. However, technological trends further increase the power density, which was expected to reach 60 kW/m^2 by the year 2010 [4]. Collectively, these factors yield an exponential increase in power needs of datacentres that is not sustainable. The focus on just speed has let other evaluation metrics go unchecked.

As a result, the total energy cost of the cooling infrastructure has increased dramatically in recent years. This evolution can be seen on Figure 1. In latest datacentres, the site infrastructure accounts for about 30% of the total energy cost. Therefore, the cooling cost is one of the major contributors of the total electricity bill of large datacentres, and another 10-15% is due to power distribution and conversion losses [6]. We can consider that one megawatt (MW) of power consumed by a supercomputer today typically requires another 0.7 MW of cooling to offset the heat generated. In 2006, datacentres in the U.S. used 59 billion KWh of electricity, costing the US $4.1 billion and 864 million metric tons of carbon dioxide (CO2) emissions; this accounted for 2% of the total USA energy budget, while it has been projected that it reached 3% by the year 2010 [7].

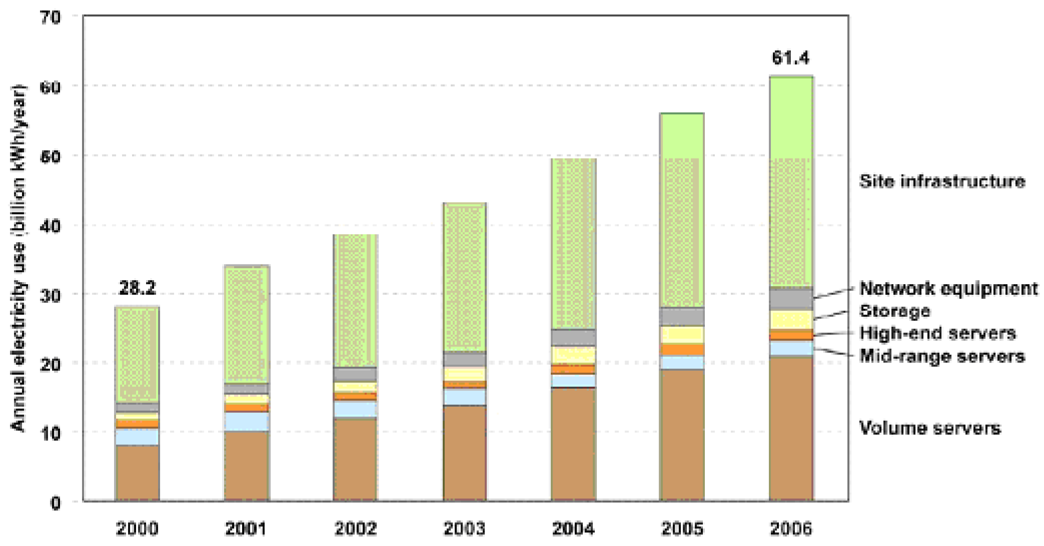

Figure 1: Electricity Use by End-Use Component. Taken from [5].

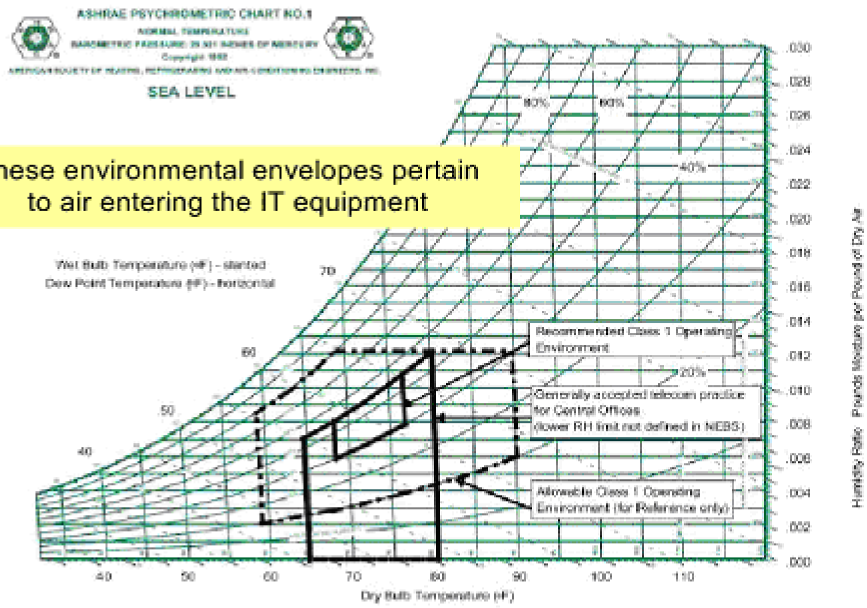

Figure 2: ASHRAE Working Environment Recommendation.

In 2010, for every $1 spent on hardware, 70 cents have been spent on power and cooling; and by 2012, for every $1 spent on hardware, $1 will be spent on power and cooling.

The advent of cloud computing has lead to an even faster growth in the number of datacentre facilities. Datacentres are beginning to appear everywhere, and they are becoming hungrier in terms of energy. This immediately translates into an urge for datacentre owners to save costs on infrastructure i.e. to make their datacentres more efficient in terms of energy and servers.

As we see next, the first urge has lead to the proliferation of green datacentres. The second one puts stress on the importance of increasing the lifetime of servers, which necessarily needs increased reliability.

2.2 The Move towards Green Datacentres

All the above mentioned factors contribute to the rise of a general concern about the high energy consumption of datacentres. Today, other metrics such as the datacentre being in the Green500 list [8] are beginning to be of importance. To be in the Green500 list [9], datacentres report the FLOPS/watt metric by measuring the average power consumption when executing the LINPACK (HPL) benchmark [10]. Also, some reference companies such as Google and IBM are already implementing measures to make their datacentres more efficient, and begin to measure the Processor Usage Efficiency (PUE) of their datacenters as a means to be aware of their efficiency. The PUE can be calculated as the total power entering by the facility divided by the power needed to run the IT equipment within it.

Also, organisms all around the world are publishing best practice documents on how to improve the energy efficiency of datacentres, and promote the creation of networks for datacentre owners and operators complying with this best practice. This is the case of the Code of Conduct on Data Centres Energy Efficiency published by the European Commission on 2008 [11]. This Code of Conduct takes into account both the IT load and the facilities load in terms of energy, and defines a metric for infrastructure efficiency, the Data centre infrastructure efficiency (DciE), which is a percentage measure of the main IT equipment energy consumption divided by the total facility energy consumption. Also, the US Department of Energy, through the Energy Star program, published a report on datacentre energy efficiency on 2007 [5], which is a compendium of data on the state-of-the-art on datacentre facilities in the US.

“Concern about the availability and security of these infrastructures, from the point of view of the backup of information, is gaining greater importance”

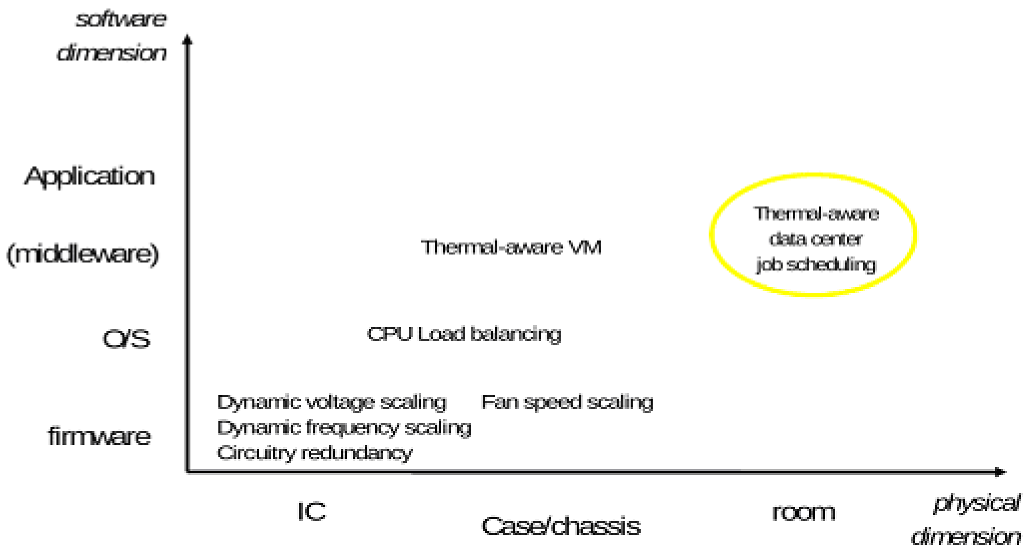

Figure 3: Energy-efficient Strategies. [Source: Impact Lab.]

2.3 The Importance of Reliability

It is important to note that energy savings can never be obtained with prejudice to reliability. Solutions to the energy problem must not suppose a decrease in the lifetime of servers or, at least, a decrease in their desired or medium lifetime. If datacentre owners need to change their equipment before they had expected to, then the costs saved in energy can be reduced, but there will be an increase due to the hardware costs.

That is the reason why some organizations such as the ASHRAE [12] publish metrics on the maximum inlet air temperature for a server, the redline temperature, as well as for the appropriate temperature and humidity conditions of the environment, so ensure that reliability is not affected. These appropriate environmental conditions can be seen on the ASHRAE psychometric chart of Figure 2. There is a lot of discussion in literature about this topic, as many authors believe that for every 10ºC increase over 21ºC in the inlet temperature will decrease the long-term reliability of electronics by 50% [13]. This belief is based on the fact that, for a constant fan speed, the CPU temperature follows the inlet temperature and so, the higher the inlet temperature, the higher the CPU temperature. ASHRAE, however, finds that these studies have not a strong basis and allows higher inlet temperatures, up to an inlet temperature of 27ºC in their 2008 recommendations. As we will see in next section, however, even though this is a key aspect, there are other issues related to reliability that can be of utmost importance, and that are sometimes overlooked.

3 Energy Efficiency in High-Performance Computing

Energy efficiency for High-Performance Computing (HPC) has traditionally centered its research in two different fields: solutions that reduce computing energy or power, some of them centered also in the reliability of circuits, and solutions to reduce the cooling infrastructure costs. Even though there should be no impediment for these two different ways to be merged into just one, the literature has mostly treated them as separate realities, or different, sometimes even independent, variables of the same problem. Only in some cases, we can find contributions that take both into account at the same time [14].

Figure 3 summarizes all the strategies regarding energy-efficiency in HPC that are explained in the next subsection.

3.1 Energy Efficient Computing and Reliability

The first solutions that take into account both reliability and energy efficiency are those that descend to the lowest level, making an approach from the circuit and chip level in the design state. This is a very common approach for reliability, as most of the effects that reduce the mean time

> “The cooling cost is one of the major contributors of the total electricity bill of large datacentres, and another 10-15% is due to power distribution and conversion losses”

> "It is important to note that energy savings can never be obtained with prejudice to reliability"

to failure, can be faced from this perspective. The effects that have been observed to have a very strong impact on reliability of systems and processor architectures are the following [15]:

- Electromigration (EM): an effect that appears due to the charge exchange between the electrons and the aluminum ions in long metal lines.
- Time-dependent dielectric-breakdown (TDDB): an important failure mechanism that models how the dielectric fails when a conductive path forms in the dielectric, shorting the anode and cathode.
- Stress migration (SM): describes the movement of metal atoms under the in?uence of mechanical-stress gradients. Stress migration may cause electrical failures due to the resistance rise associated with the void formation.
- Thermal cycling (TC): this effect produces a permanent damage that accumulates each time the device undergoes a normal power-up and power-down cycle.

Even though all the effects are present in datacentres, the most common are EM, caused when hot spots appear in processors, or TC, as a result of uneven load balancing in servers.

The reduction of hot-spots, apart from increasing reliability, lowers the cooling effort because not so high temperatures are achieved in the chip. That is why many approaches in literature are focused on the reduction of hot-spots. This is the case with all the literature on temperature-aware floor-planning of cores (especially in the world of MPSoCs) [16][17][18], which is devoted to getting the optimum floorplan that reduces the hot-spots. Also, hot-spots can be reduced by means of temperature-aware task allocation and scheduling algorithms [19].

Even though, at a first glance, the world of MPSoCs seems far apart from the world of datacentres, the usage of embedded systems in HPC to allocate tasks that do not have hard time constraints is becoming more important by the day. Some papers focus on trying to find the most energy-efficient datacentre blocks between machines from the embedded, mobile, desktop and server spaces [20].

Going up in the level of abstraction we find some common approaches for thermal-aware task scheduling at the operative system level, such as heat balancing or deferred execution of hot jobs [21]. This is the case with temperature-aware Greedy algorithms that prevents hot-spots and imbalanced temperature distributions, while ensuring that all servers inlet temperatures are greater than a threshold, increasing the priority to retirement when the condition is not accomplished [4].

However, it is at the server level where the most popular measures to reduce power consumption without impact on performance are found. These measures are those based on Dynamic Voltage Frequency Scaling (DVFS) [22] which is a power-aware algorithm that automatically and transparently adapts the voltage and frequency settings of servers, Vary-On Vary-Off (VOVF). and all techniques based on switching off idle machines, in order to save the costs of keeping on under-saturated machines [23].

At the highest level, the last two techniques can be used alone or jointly with power-aware scheduling strategies [24], even taking into account system load considerations or task characterization [25]. Efficient job allocation [26] by means of adaptive power-aware scheduling algorithms help to reduce peak and average power consumption.

Less common approaches try to reduce power by means of fans, networking or IO operations.

The newest approaches are beginning to take into account some capabilities that are mostly used in Cloud Computing, as we will see next, such as virtualization solving multi-objective scheduling to minimize energy in virtualized datacenters [27]. Virtualization enables the consolidation of multiple workloads in a smaller number of machines. So, even though it incurs some additional overheads, because of virtual machine creation and migration dynamic job scheduling policies still mean considerable energy reductions.

3.2 The Cooling Perspective

Energy efficiency from the cooling perspective can exploit two different facts.

On the one hand, cooling efficiency can take into account that there are some physical locations in the datacentre which are more efficient to cool than others. If heavier workloads are placed in the easier-to-cool places, scheduling is more efficient in terms of energy [7]. A refined version would be the one that studies which are the steady state hot-spots and cold-spots of the datacentre, and uses them to generate temperature-aware workload placement [13].

On the other hand, some approaches try to maximize the temperature of the air supplied by the air conditioning units, or to minimize the heat recirculation of the datacentre so that cooling is more efficient e.g. in [28] less tasks are assigned to chassis that are known to have more recirculation. Some algorithms are known to work well for these conditions and generate an energy efficient job scheduling. Some of the most common algorithms are [29]:

- Uniform Outlet Profile: that tries to make uniform the outlet temperatures of all servers.
- Minimal Computing Energy: which generates schedules that spend the minimum energy.
- Coolest Inlet: which distributes the workload first to the servers with lowest inlet temperature [30].

Finally, some task allocation algorithms try to balance load, migrating tasks from thermally saturated cores to cooler neighbors [31].

Both approaches are, after all, based on taking into account the thermodynamic effects of the server room. To prove the abovementioned hypotheses, it is very common to use Computational Fluid Dynamics software, such as Mentor Graphics Flovent [32]. This software tool computes the specific performance of data rooms, and allows researchers to prove the results provided by the algorithms. Typically the datacentre is considered a warehouse-sized facility with several rows of server racks. The room is usually considered to be of the hot-aisle/cold-aisle style, that is, each row is sandwiched between a hot aisle and a cold aisle. The cold air is supplied by the air conditioning units through perforated tiles in the floor underneath the cold aisles. The servers suck the cold air coming from the cold aisle into the rack using chassis fans, and push out the heated air to the hot aisle. The hot air is extracted from the air-conditioning units at the room ceiling.

4 The Contribution of Cloud Computing

The move towards cloud computing has given a new vision to the energy efficiency problem. The key current technology for energy-efficient operation of servers in cloud datacentres is virtualization. The cloud datacentre is, by definition, heterogeneous. Heterogeneity might be overlooked in datacentres for HPC, even though the real datacentres are almost always heterogeneous, due to the different versions of their hardware, but it must be taken into account in Cloud Computing datacentres. Virtualization is useful in the sense that it can abstract the heterogeneity of the servers and thus make allocation and migration easier. This however, cannot be by itself the only practice applied. Some solutions for Virtual-Machine based datacentres propose to investigate the resource demands characterization of each workload and the live migration of virtual machines without QoS degradation [33], finding the minimum energy allocation of workloads to servers, using heuristics that maximize the sum of Euclidean distances for current allocations or avoiding bottlenecks so that machines are never idle [34]. Cloud computing provides opportunities for the usage of turn off policies and optimization, or even re-factory, of network protocols in a way that enhances the energy-efficient operation of the network elements. As cloud computing is becoming of significant importance, the amount of data that is transferred over the Internet is increasing exponentially. In order to create green cloud computing datacentres, content replication and dissemination algorithms will then need to consider energy as a key parameter of optimal operation [35].

The problem of Best-Fitting, that is, the problem of assigning each VM to the resource where it will best behave from the energy-efficiency point of view, as well as the policies for migrating that VM somewhere else, is comprehensive in Cloud Computing. The most common algorithms to approach this problem are [36][37]:

- Minimization of Migrations (MM): which consists of migrating the least number of VMs to minimize migration overhead as this overhead has an associated energy consumption.
- Highest Potential Growth (HPG): which consists of migrating VMs that have the lowest usage of CPU in order to minimize total potential increase of the utilization and SLA violation.
- Random Choice (RC): which consists of choosing the necessary number of VMs by picking them according to a uniformly distributed random variable.
- Modified Best Fit Decreasing (MBFD): this algorithm consists of sorting all VMs in decreasing order depending on their current CPU utilizations, and allocating each VM to a host that provides the least increase in power consumption due to this allocation (choosing the most power-efficient nodes first).

As it can be seen, the need for cloud computing to become green has lead to a proliferation of algorithms that take into account virtualization, heterogeneity and high network demands. This opens a new path in research, which could feed back some of the cloud solutions to improve the energy efficiency of HPC datacentres.

5 Real Solutions in Today's Datacentres

First of all, it must be taken into account that, from the datacentre owner's point of view, their facilities are something more than just air conditioning units, racks and servers. Datacentres have many security measures such as fire detection alarms, fire suppression, smoke dampers, fresh air supply and emergency off buttons. They are also real production and working environments that cannot assume to integrate any new measure that can put at a stake their availability or reliability. All research solutions that aspire to be brought to market must keep in mind the security restrictions of datacentres and the need for the machines to be always ready to absorb the changing loads of a datacentre, ensuring that the stability of the machines will never be at stake.

There are of course some measures that have already been applied to datacentres. Most of them, however, are more related with best practices than with innovative techniques of workload scheduling. It is true that datacentres, such as the those with the BlueGene computers from IBM,

> "An important gap exists between the contributions of researchers and the products offered by industry"

which are at the top of the Green500 list, or Google, are beginning to constantly measure their PUE, and have realized the benefits of having good airflow management and better isolation between hot and cold aisles. The trend of keeping the inlet temperature "at the lowest possible", however, continues to be a common practice, even though this is not a recommendation anymore and does not serve the cause of energy efficiency nor the reliability of servers.

Only some of the most innovative companies, such as Google, are applying measures to make their datacentres Green so that they can save energy costs. Google announced in 2008 that their Container Datacentre [38] had an effective PUE of 1.25. This high efficiency is reached on the one hand by the use of heat exchangers, optimized power distributions, that increase the efficiency of transformers and UPS, and free cooling. On the other hand, Google pays close attention to the airflow management: their datacentres try to eliminate hot/cold air mixing by isolating the aisles. Their machines work in cold aisles with temperatures of 27ºC, that is, at the limits recommended by the ASHRAE organization in 2008.

6 Conclusions

The reality of today's datacentre seems, at a first glance, far apart from the research on the field. The solutions given by University departments are far away both from commercial datacentres and from the solutions sold by industry.

Most contributions assume unreal datacentres; smaller rooms than the real ones with severe hot/cold air mixing and recirculation. Even though that is the case with many datacentres, there are commercial solutions based on better airflow isolation that solve the problem. However, the really important contributions, that is regulation over air conditioning units and energy-efficient scheduling, has not yet been integrated in real datacentres. Some trials have been carried out in experimental datacentres, but these solutions have not yet been put into industrial use.

The most energy efficient datacentres – those that are in the first positions of Green500 list – do not use automatic energy-aware scheduling algorithms nor automatic air conditioning supply temperature control. They just follow the best practices and recommendations of the abovementioned institutions.

An important gap exists between the contributions of researchers and the products offered by industry. Even though some solutions seem ready for the market, innovation in this field has lots of barriers when arriving to real production environments. Companies must be very positive about the strength and reliability of the proposed solutions before they allow testing in their datacentres, as unstable solutions could lead to severe economic losses for datacentre owners.

Acknowledgements

This work was partially funded by the Spanish Ministry of Science and Innovation through the Secretariat of State for Research, under Research Grant AMILCAR TEC2009-14595-C02-01, through the General Secretariat of Innovation under Research Grant P8/08 within the National Plan for Scientific Research, Development and Technological Innovation 2008-2011 and the Campus of International Excellence (CEI) of Moncloa, under Research Grant of the Program for Attracting Talent (PICATA).

References

[1] TOP500 Supercomputing Sites, 2009. [Retrieved October 25, 2011, from <http://top500.org/>.]

[2] W. Feng, K. Cameron. "The Green500 List: Encouraging Sustainable Supercomputing". Computer, 40(12), pp. 50-55, 2007.

[3] J.M. Kaplan, W. Forrest, N. Kindler. "Revolutionizing Data Center Energy Efficiency". McKinsey & Company, July 2008.

[4] E. Pakbaznia, M. Pedram. "Minimizing data center cooling and server power costs". Proceedings of the 14th ACM/IEEE international symposium on Low power electronics and design - ISLPED '09 (p. 145). New York, New York, USA: ACM Press, 2009.

[5] U. D. Of Energy. "Report to Congress on Server and Datacentre Energy Efficiency". Public Law pp. 109-431, 2007.

[6] C. Bash, G. Forman. "Cool job allocation: Measuring the power savings of placing jobs at cooling-efficient locations in the datacentre". 2007 USENIX Annual Technical Conference on Proceedings of the USENIX Annual Technical Conference (pp. 17-22). USENIX Association, 2007.

[7] T. Mukherjee, A. Banerjee, G. Varsamopoulos, S.K.S. Gupta, S Rungta. "Spatio-temporal thermal-aware job scheduling to minimize energy consumption in virtualized heterogeneous datacentres". Computer Networks, 53(17), pp. 2888-2904, 209.

[8] The Green500 List. Environmentally Responsible Supercomputing. [Retrieved October 25, 2011, from <http://www.green500.org/>.]

[9] B. Subramaniam, W. Feng. "Understanding Power Measurement Implications in the Green500 List". 2010 IEEE/ACM Int'l Conference on Green Computing and Communications & Int'l Conference on Cyber, Physical and Social Computing (pp. 245-251). IEEE, 2010.

[10] Linpack HPL Benchmark for HPC. [Retrieved October 10, 2011, from <http://www.netlib.org/benchmark/hpl/>.]

[11] European Comission, European Union's Joint Research Centre (JRC). "Code of Conduct on Data Centres Energy Efficiency Version 1.0". October, pp.1-20, 2008.

[12] American Society of Heating, Refrigerating and Air-Conditioning Engineers. "ASHRAE Environmental Guidelines for Datacom Equipment - Expanding the Recommend Environmental Envelope", 2008.

[13] J. Moore, J. Chase, P. Ranganathan, R. Sharma. "Making scheduling 'cool': Temperature-aware workload placement in datacentres". USENIX Association Proceedings of the General Track: 2005 UNENIX Annual

Technical Conference, pp. 61-74, 2005.

[14] A. Banerjee, T. Mukherjee, G. Varsamopoulos, E.K.S Gupta. Integrating Cooling "Awareness with Thermal Aware Workload Placement for HPC Datacentres". Impact Lab. Arizona State University, 2011.

[15] D. Atienza, G. De Micheli, L. Benini, J.L. Ayala, P.G.D. Valle, M. DeBole, V. Narayanan. "Reliability-aware design for nanometer-scale devices". Proceedings of the 2008 Asia and South Pacific Design Automation Conference (pp. 549–554). IEEE Computer Society Press, 2008.

[16] Y. Han, I. Koren, C. Moritz. "Temperature aware floorplanning". Proc of the Second Workshop on TemperatureAware Computer Systems, 2005.

[17] Y. Han, I. Koren. "Simulated Annealing Based Temperature Aware Floorplanning". Journal of Low Power Electronics, 3(2), pp. 141-155, 2007.

[18] W. Hung, Y. Xie, N. Vijaykrishnan, C. Addo-Quaye. "Thermal-aware floorplanning using genetic algorithms". International Symposium on Quality Electronic Design, 2005.

[19] Y. Xie, W.L. Hung. "Temperature-aware task allocation and scheduling for embedded multiprocessor systems-on-chip (MPSoC) design". The Journal of VLSI Signal Processing, 45(3), pp. 177–189. Springer, 2006.

[20] L. Keys, S. Rivoire, J.D. Davis. "The Search for Energy-Efficient Building Blocks for the Datacentre". Workshop on Energy-Efficient Design, 2007.

[21] J. Choi, C.Y. Cher, H. Franke, H. Hamann, A. Weger, P. Bose. "Thermal-aware task scheduling at the system software level". Proceedings of the 2007 international symposium on Low power electronics and design (pp. 213–218). ACM New York, NY, USA, 2007.

[22] C-H.Hsu, W-C. Feng. "A Power-Aware Run-Time System for High-Performance Computing". ACMIEEE SC 2005 Conference SC05, (November), 1-1. IEEE, 2005.

[23] X. Zheng, Y. Cai. "Markov Model Based Power Management in Server Clusters". International Conference on Green Computing and Communications (pp. 96-102). IEEE, 2010.

[24] X. Zhu, C., He, Y. Bi, D. Qiu. "Towards Adaptive Power-Aware Scheduling for Real-Time Tasks on DVS-Enabled Heterogeneous Clusters". 2010 IEEE/ACM Int'l Conference on Green Computing and Communications & Int'l Conference on Cyber, Physical and Social Computing (pp. 117-124). IEEE, 2010.

[25] W. Felter, K. Rajamani, T. Keller, C. Rusu. "A performance-conserving approach for reducing peak power consumption in server systems". Proceedings of the 19th annual international conference on Supercomputing ICS 05, pp. 293-302. ACM Press, 2005.

[26] C. Bodenstein, G. Schryen, D. Neumann. "Reducing Datacentre Energy Usage Through Efficient Job Allocation". ECIS 2011 Proceedings. Brill, K. G. (2007).

[27] J. Xu, J.A.B. Fortes. "Multi-Objective Virtual Machine Placement in Virtualized Datacentre Environments". 2010 IEEE/ACM Int'l Conference on Green Computing and Communications & Int'l Conference on Cyber, Physical and Social Computing (pp. 179-188), 2010.

[28] Q. Tang, E.K.S. Gupta, G. Varsamopoulos. "Thermal-aware task scheduling for datacentres through minimizing heat recirculation". In IEEE Cluster, 2007.

[29] S.K.S. Gupta, G. Varsamopoulos. "Energy-Efficient Thermal-Aware Task Scheduling for Homogeneous High-Performance Computing Datacentres: A Cyber-Physical Approach". IEEE Transactions on Parallel and Distributed Systems, 19(11), pp. 1458-1472, 2008.

[30] G. Varsamopoulos, A. Banerjee, S.K.S. Gupta. "Energy efficiency of thermal-aware job scheduling algorithms under various cooling models". Contemporary Computing, pp. 568–580. Springer, 2009.

[31] G. Wu, Z. Xu, Q. Xia, J. Ren, F. Xia. "Task Allocation and Migration Algorithm for Temperature-Constrained Real-Time Multi-Core Systems". 2010 IEEE/ACM Int'l Conference on Green Computing and Communications & Int'l Conference on Cyber, Physical and Social Computing (pp. 189-196). IEEE, 2010.

[32] Flomerics Ltd. "Flovent version 2.1," Hampton Court, Surrey, KT8 9HH, England, 1999. Available at <http://www.flomerics.com/>.

[33] K. Ye, D. Huang, X. Jiang, H. Chen, S. Wu. "Virtual Machine Based Energy-Efficient Datacentre Architecture for Cloud Computing: A Performance Perspective". 2010 IEEE/ACM Int'l Conference on Green Computing and Communications & Int'l Conference on Cyber, Physical and Social Computing, pp. 171-178. IEEE, 2010.

[34] S. Srikantaiah, A. Kansal, F. Zhao. "Energy aware consolidation for cloud computing", 2008.

[35] A. Berl, E. Gelenbe, M. Di Girolamo, G. Giuliani, H. De Meer, M.Q. Dang, K. Pentikousis. "Energy-Efficient Cloud Computing". The Computer Journal, 53(7), pp. 1045-1051, 2009.

[36] A. Beloglazov, J. Abawajy, R. Buyya. "Energy-aware resource allocation heuristics for efficient management of datacentres for Cloud computing". Future Generation Computer Systems, 2011.

[37] R. Buyya, A. Beloglazov, J.Abawajy (n.d.)." Energy-Efficient Management of Datacentre Resources for Cloud Computing: A Vision, Architectural Elements, and Open Challenges".

[38] Official Google Blog: Designing lean, green, energy-saving machines. (n.d.). [Retrieved October 7, 2011, from <http://googleblog.blogspot.com/2009/04/designing-lean-green-energy-saving.html>.]

