

From the Origins of Performance Evaluation to New Green ICT Performance Engineering

Carlos Juiz and Ramon Puigjaner

Universitat de les Illes Balears
Carretera de Valldemossa, km 7.5
07122 PALMA, Spain
cjuiz@uib.es, putxi@uib.cat

Abstract. This paper intends to present an overview of the evolution of performance evaluation since its first steps the Erlang works for modelling telephone networks, based on simple queues until the present current challenges in Green ICT that will require the development of new paradigms and mathematical tools, and rapidly passing across the modelling works of Khintchine and Pollaczek; Jackson; Baskett, Chandy, Muntz and Palacios; Buzen; Reiser and Lavenberg; and many others, and benchmarking standards that have produced solutions to the problems appearing in these hundred of years. Finally, we analyze some of the challenges of computer performance evaluation appearing today, mainly those related to the energy consumption and sustainability, globally known as Green ICT.

Keywords: Performance evaluation of telephony, computer systems and communication networks, Performance modeling, Queuing theory, Queuing network theory, Simulation, Benchmarking, Green ICT.

1 Introduction

It is well known that you cannot manage what you cannot measure and also that before installing a complex system it is better to have an estimation of what will be its performance behaviour. In the domain of information and telecommunication technologies (ICT), the first case of application of the rules of thumb we mentioned was based on the Erlang works on queuing theory to plan the capacity of the telephone system [1], [2], [3]. Since then a lot of techniques and tools have been developed for predicting the behaviour of ICT systems. These techniques have evolved to be able to tackle the new challenges that have appeared during the last hundred years.

This paper is organized as follows: in the next section the first works of Erlang applied to telephone systems are presented. Section 3 rapidly explores the improvements of the queuing theory experienced until the moment in which computer appears. Section 4 reviews the different models and benchmarks developed to predict the performance of computer systems. In section 5 models and benchmarks developed for communications systems are presented. Section 6 presents the evolution of operational drivers of success for systems and how Performance Engineering is

evolving through these drivers. Section 7 presents the challenges of analyzing the performance of new types of ICT systems or more precisely the challenges of new concepts in performance evaluation. And finally section 8 concludes.

2 Erlang Works

Even if the works of **Andrei Andreyevich Markov** obtained his results on Markov chains (refined later by Norbert Wiener) before those of Agner Karup Erlang, we prefer to start our history with Erlang because his works are more related to ICT systems. Erlang was a Danish mathematician that worked for long time at the Copenhagen Telephone Company where he had the opportunity to analyze the phenomena related to the new invention that was the telephone. Among his works they are specially important the analysis allowing to represent the telephone communications like a Poisson distribution [1] and the determination of how many circuits were needed to provide an acceptable telephone service (B and C Erlang formulae) and how many telephone operators were needed to handle a given volume of calls [2], [3]. It is necessary to remember that at the beginning of the twentieth century human operators and cord boards were used to switch the telephone calls by means of jack plugs. The main limitation of the Erlang works is the assumption that both the service time and the interarrival time should be exponentially distributed. These results were intensively used for more than sixty years.

However, Erlang was not a pure researcher but a hands-on one. To verify his assumptions he did not refuse to conduct measurements that obliged him to climb into street manholes. He can be considered as the father of the queuing theory and the performance evaluation of ICT systems. He was also an expert in the history and calculation of the numerical tables of mathematical functions, particularly logarithms. He devised new calculation methods for certain forms of tables.

3 Other Advances before the Computers

The next important result in queuing theory came from two mathematicians that simultaneously break one of the limitations of the Erlang assumptions: the service time could have any type of distribution. This result was obtained in 1930 by the Austrian-French Félix Pollaczek and the Russian Aleksandr Yakovlevich Khinchin working independently one of the other. In this case the result was not induced by a technical need but was of academic interest.

Other interesting results, useful in some computer models, were those obtained by Alan Cobham who analyzed the behaviour of a queuing system when there are customers with different priorities. His initial work can be found in [4].

4 Performance Evaluation of Computers

The need of predicting the performance of computers obliged scientists and engineers to develop new techniques and tools adapted to the new challenges.

4.1 Modelling

Although the initial step for tackling the problem of computer system performance modelling was initially thought for modelling a computer system performance, the pioneering works of Jackson [5] and Gordon and Newell [6] with the algorithmic complement of Buzen [7] prepared the arrival of the work of Baskett, Chandy, Muntz and Palacios [8] on product form networks. This work established a complete basis for modelling the performance of a computer system allowing an analytical process, even if the possibility of analytical processing was restricted by an important set of conditions. This work was completed by the proposal of convenient algorithms like those of Reiser and Kobayashi [9] and Reiser and Lavenberg [10].

In parallel with these works, Buzen y Denning [11], [12] developed a different approach for tackling the same problem and arrived to similar numerical results; the operational analysis. This approach has a more physical basis without the strong mathematical and statistical conditions of the queuing network approach.

In order to facilitate the use of these results to engineers and researchers, several tools were developed. Among them we can find BEST-1 developed by BGS, the company created by J.P. Buzen, RESQ developed by IBM and QNAP2 developed by INRIA.

In order to free the modelling process of the constraints of the product form networks, a number of approximated methods were developed, like those of Courtois [13], Marie [14], Gelenbe and Mitrani [15] and many others. Some of these approximated algorithms were included in some of the previous tools.

Some of the above mentioned tools also included simulation capabilities to analyze queuing network models not able to be processed by any other technique.

4.2 Benchmarking

However in many cases computer customers need to compare the computing capacity of different systems. Initially this capacity was evaluated simply by the execution time of a typical instruction, habitually the addition instruction. With the increasing complexity of computer systems (clever architectures, sophisticated operating systems, etc.) this simple comparison was no more valid. Computer customers proposed the execution of a sample of their real workload on the different systems they had to compare. Sophisticated techniques were developed to characterise the system workload by a reduced number of programmes extracted from this workload.

The next step was to standardise these sets of programmes depending on the type of workload we were interested in. The definition of these standards was reached by consensus like in the case of LINPACK [16], a collection of Fortran subroutines that analyze and solve linear equations and linear least-squares problems for the scientific computation or established by organizations participated by computer manufacturers and customers, like in the case of SPEC [17].

The first benchmarks proposed by SPEC were devoted to batch environments, then the target was a conversational system and currently cover most of the typical working environments of computer systems. An aspect particularly interesting is the possibility of scaling the benchmarks in order to adapt the size to the particular needs of the customer. Also SPEC organization publishes results of its benchmarks with different sizes running over a variety of systems.

5 Performance Evaluation of Networks

When computers were no more able to work isolated the need of connection initially with terminals and then with other computers appears. The complexity of the systems increased and consequently the complexity of the tools and techniques used to evaluate their performance.

5.1 Modelling

When the computers ran in an isolated way it was quite easy to build analytical models able to be processed with the help of appropriated tools. However, with the increasing of the networked systems complexity (based on wired or wireless LANs, or in ad-hoc networks, or in sensor networks, etc.), frequently it was very hard to find an enough accurate model for representing such systems and able to find the solution analytically. Fortunately, the increasing computer speed allowed building simulation models representing with enough accuracy the behaviour of the corresponding systems and having the possibility of obtaining results with small enough confidence intervals in computing times reasonably short.

Among the different simulation engines there two that are used very frequently: OPNET [18] and ns-2 [19].

5.2 Benchmarking

In terms of standard benchmarks, and restricting our comments to SPEC [17], the variety of benchmarks proposed by this institution has been increased to cover the most typical systems including a network in their structure. To cover these needs benchmarks for analyzing conversational systems, transactional systems, web systems (an e-commerce sales systems), etc. have been developed and proposed to the community of computer systems users.

6 New Performance Challenges of ICT Systems

The classical drivers of operational success at ICT have been always: Performance, Dependability, Quality, Flexibility and Cost [59]. Some of them seem to be clearly non-functional and other functional, but all five are closely related. In fact, the cost has the key of selecting particular solutions at datacenters for the other four operational drivers. Lately, other two drivers have been emerged as important as the first four: Security (some authors included it into Dependability) and Energy (classically included it in Cost). In Sections 2-5 we have overviewed the evolution, just considering Performance, then considering Performance and Quality or Performance and Dependability, and other combinations of operational drivers for Performance Engineering and Performance Evaluation of software, hardware, networks, etc. In last ten years, a lot of research effort has been put on Performance and Security issues, but the issue of this paper is about the new opportunities in the combination of Performance and Energy efficiency, which is part of Green ICT arena. That is, most of old studies in Performance Evaluation were done with a single driver

in mind: Performance, but nowadays, there is a shift between this single objective to the consideration of multiple, particularly, Performance and Energy.

Turning to the drivers of success, it is clear that taking only one of them usually produce the detriment of some of the others. For example, maximizing the performance may not guarantee quality of service or even do more inflexible a particular solution, while maximizing security can reduce yields by overexposure to controls, and these situations always increase the cost. Consequently, the challenge of the new performance engineers is to find a balance among these factors. Similarly, one of the challenges is to find the Green ICT energy efficiency acceptable performance, i.e. the energy cost control while maintaining the performance necessary for operation. In the next section, we try to explain the factors that can influence the performance engineering and energy, and the opportunities offered by these factors for research.

7 Green ICT and Performance

The prevailing international scientific opinion on climate change is that human activities resulted in substantial global warming from the mid-20th century, and that continued growth in greenhouse gas concentrations [48] caused by human-induced emissions would generate high risks of dangerous climate change [32]. Even this debate is controversial, what it is clear is that some human activities, as the use of ICT, are producing huge greenhouse gases.

With the overall improvement in the quality of connections and the spread of the wireless Internet, the web applications, easily updatable and often “in the cloud and free”, are gaining ground rapidly to traditional software. But not only distributed applications of cloud computing are growing, traditional web applications based on client-server paradigm are also growing, despite the alleged gratuity of some popular web applications.

Data center power consumption continues to grow at an alarming pace [37, 28]. There are five distinct industry segments accounting for most of a data center’s energy-efficiency: (1) servers and storage systems, (2) power conditioning equipment, (3) cooling and humidification systems, (4) networking equipment, and (5) lighting/physical security. Therefore, sustainable data centers require the multidisciplinary collaboration of mechanical engineers, electrical engineers, computer scientists, and others. Particularly, the hardware infrastructure of the data center consists of hundreds (or thousands) of servers hosting revenue-generating services, interconnected with each other and the outside world via networking equipment, and relying on storage devices for persistent data. These hardware elements are managed by a datacenter-wide software stack that spans the platforms and virtualization layers [50].

On one hand, data centers that host web applications and user files are huge warehouses with a lot of servers, additional electronic equipment and complex cooling systems. While computers, mobile phones and telecommunications networks are today primarily responsible for energy consumption, energy spending caused by data centers that maintain the quality of internet services in operation is growing

faster. But giving more capacity to data centers usually means more servers and consequently all five factors mentioned above consume more energy [51].

On the other hand, Performance engineering is the set of roles, skills, activities, practices, tools, and deliverables applied at every phase of the Systems Development Lifecycle, which ensures that a solution will be designed and implemented to meet performance requirements [44]. The increasing number of Internet users and systems connected to it, especially due to the popularization of applications and Web-based services, produces continuously capacity problems. In these growing web systems, if no action is taken, a loss in quality of service is produced and, ultimately, worsens the performance of any web application. Additionally, this growth of web systems, may be in a slightly irregular shape and unplanned in most cases, due to time-to-market pressures [46]. Variables that could be affected by these changes in system workload are those perceived by the user and that constitute the quality of service (Quality of Service, QoS, among which is singular importance, the response time) as those need to know the server's responsibility to provide quality of service desired (which shall include the use, overhead and the throughput) [56]. The performance improvement of these services can be done at different layers, from server side to applications, from network protocol to routing, etc. [45], [47]. When even the techniques of performance and modelling of web-based computer systems and distributed has not reached a sufficiently scientific mature, a new challenge directly related to this branch of research emerges: the relationship between performance and energy cost is i.e., the energy efficiency of systems; this section is aimed by these combination of two topics and the future challenges for performance engineers due to this combination.

7.1 State of the Art

The solutions to achieve better energy efficiency in sharing distributed resources will address different topics. The first is at the computer architecture level, techniques have been proposed to reduce processor energy usage through multi-core processors [34], designing energy proportional hardware [20], and increasing the number of performance and sleep states [41]. Benchmarking techniques to quantify the energy costs of different computer architectures have been proposed [48].

Even though, the current possibilities of on-the-fly adaptation of the underlying hardware behaviour will be studied: for instance even it is possible to decrease the voltage of the processors to slow them down and save energy, there will be the need to characterize the electrical consumption of the systems, as a function of the workload of the system. This implies first to have a means to measure or estimate the energy consumption volume (by actual power-consumption measurement equipments), and then to model the behaviour of the computations related to the obtained performances. However, the price to be paid is reducing the system components longevity. Actions have to be taken at different levels of decision to improve the energy-efficiency of the distributed web servers:

- At the networking level, the idea would be to decrease the overall usage of the network. For instance, moving large amount of data will operate on routes consuming less energy and the placement of the data (or the replication of the data) will be optimized to minimize the equipment usage; complementarily, approaches to minimizing the time to transfer the data, and to reducing the data to

be transmitted (e.g., by using forecasting techniques) will be investigated.

The performance forecasting for several network layers [31], [39] should provide some skills performance and power tools development in order to know how applications are using the network.

- At the server level, in order to maximize energy efficiency, whether in a single system or over an ensemble of systems, users and data center operators need to understand the relationship between resource usage and system-level power consumption. This understanding enables such optimizations as consolidating workloads on as few machines as possible and turning others off [22], [25], [35], since current hardware is highly inefficient at low utilization [20]. While several types of full-system power models have been proposed, often in the context of enabling a particular optimization, they have not been systematically compared over a variety of software workloads and hardware platforms.

Performance engineering will contribute to these efforts building several power models for servers executing web different workloads in extensive studies including regression models and also simulation. Other techniques include multi-server power coordination [43], hard disk spin-down [40], synchronization aware multithreading [42], and compiler driven optimizations [49].

Another interesting issue is server virtualization, where hardware elements are managed by a datacenter-wide software stack that spans the platforms and virtualization layers. However, adding a new software layer also adds additional overhead and risk of over utilization of hardware with clear consequences in energy but also savings.

Performance engineers will also study how performance and energy varies with virtualization of servers in comparison with the non-virtual ones.

- At the application level, context awareness of the applications should be favored: it aims at developing applications that are aware of the power state of the computers and the currently set power policy, such that they are tuned appropriately, and respond to changes dynamically [38]. Developing energy-aware data mining techniques will support energy efficiency at the applications level. For instance, clustering/classifying of data sources, identifying usage patterns and cluster users will help to decrease the energy consumed in the network. This last level is very attractive to performance research for several reasons. First, this layer has the most information on the actual user impact of performance and energy tradeoffs, enabling more aggressive performance sacrifice in unimportant areas compared to lower layer techniques. For instance, an application may throttle processor frequency to the minimum value required to satisfy user perceptible delay behavior. Second, application specific optimizations can be made at this layer such as changing the algorithm used, accuracy of computation (e.g. changing from double precision to single), or quality of service provided [29]. Third, energy usage at the application layer may be made dynamic [27]. For instance, an application hosted in a data center may decide to turn off certain low utility features if the energy budget is being exceeded, and an application on a mobile device may reduce its display quality [30] when battery is low. This is different from system layer techniques that may have to throttle the throughput resulting in users being denied service.

7.2 Some Green ICT Research Challenges for Performance Engineering

Many of the current green performance activities represent isolated optimizations; others attempt to standardize approaches and solutions and to promote a common understanding of the issues and complexities. While recognizing their vital importance, the literature is replete with such solutions and experiences. The new challenges will be how to reduce power consumption in a collaborative manner not only as individual efforts on devices but holding an optimal quality of service.

These new performance challenges will propose realistic energy-efficient alternate models and also study new solutions for benchmarks. While much effort is nowadays put into hardware specific solutions to lower energy consumptions, the need for a complementary approach is necessary from the application viewpoint. Performance Engineering will characterize the energy consumption and energy efficiencies of applications at servers. Then based on the current hardware adaptation possibilities and interconnection networks, performance engineers will characterize the trade-off between energy savings and functional and non-functional parameters, particularly the web system performance. Most of the current effort consists of building performance and power models for servers from standard benchmarking, subsequently deriving new simulation engines to derive performance and energy results and finally propose a new benchmark models for servers taking into account energy issues.

For this reason, we propose an incomplete list of Green ICT performance topics putting together computing technologies, networking, optimization, and their performance evaluation and power issues. Our proposed research opportunities list focuses on several integrated energy and ICT related issues, from simulation to software applications, from networking to data centres, from performance engineering to consumption models, etc. trying to transit from research on only ICT functional issues to green ICT considerations, together.

The current open list of research topics can be started with the first volume of results in COST IC0804 [58], where there are included the results of this European action. However, our position is that there are research issues very appealing for the performance and power research advance. Here there is our open and incomplete list:

- Energy consumption/performance models: models and tools where system administrator concerns about energy-efficiency of ICT usage are considered together with performance.
- New simulation and benchmarking and monitoring engines considering devices, networks and their corresponding energy consumption
- Energy Efficient Computing Centres: this will be addressed at different levels, i.e. at a data centre level, workload and server management, respectively.
- Energy Efficient Networks: at multiple levels, including networking devices, communication and control protocols, network architectures design, and network performance.
- Green Software Design: considering the compiler level on algorithmic and data structures optimizations for grid systems. Also, at application level to investigate what are the key aspects in the applications that specifies the performance requirements/power consumption that user expects from those services

- Design for Energy Awareness: empowering social in their use of products and services by increasing visibility of energy choices during operation.
- ICT Management Green Metrics: indicators of performance/energy from data centre monitoring to ICT governance/management
- Power-aware middleware for data centres: more intelligent power management for virtualised machines, power-aware virtualisation and benchmarking.
- Power-efficient routing: design of packet classifiers for high speed and low power consumption.
- Energy-efficient bandwidth allocation: making minimum consumption of energy resources in network equipment.
- Energy and distributed computing: placement of mobile agents, workload placement, virtual machine migration, P2P resource allocation, etc, providing means to develop energy-aware allocation strategies.
- Eco Data Base Management Systems: managing the energy consumed during query processing and considering query aggregation in workload.
- Energy-Efficient algorithms: Dynamic Speed scaling and power-down client integration but holding resilience of systems.
- Dynamic performance management: programming models for the cloud rely on the use of tasks scheduled across large data centres, taking advantage of the parallelism/performance and power.
- Etc.

For sure that there are many more, in fact, all performance engineering issues are directly or indirectly related to power consumption, so that we may select almost every performance issue to be studied from the energy-efficiency viewpoint. Thus, we may review the performance engineering research from the very beginning but now considering power consumption as a primary factor for the new Systems Development Lifecycle.

8 Conclusions

Performance evaluation concept has evolved along with the changes experienced by the changes in the technology: hardware, software and networks, and in the goals of the end users with respect to the systems they were using.

We have overviewed the evolution, just considering Performance evaluation, then considering other engineering factors as Performance and Quality or Performance and Dependability, and other combinations of operational drivers for Performance Engineering of software, hardware, networks, etc. In last ten years, a lot of research effort has been put on Performance and these new issues, but the topic of this paper is about the new opportunities in the combination of Performance and Energy efficiency, which is part of Green ICT arena. That is, most of old studies in Performance Evaluation were done with a single driver in mind: Performance but nowadays, there is a shift between this single objective to the consideration of multiple, particularly, Performance and Energy. We have seen how the tools and techniques used to evaluate or predict the computer performance have evolved and the new challenges derived from Green ICT will oblige to find new tools and techniques addressed to overcome the current challenges as they are suggested in the last part of this paper.

Acknowledgements. This work has been partially funded by the Spanish Ministry of Science and Innovation and the FEDER funds, both in the framework of the projects TIN2007-60440 and TIN2009-11711.

References

1. Erlang, A.K.: The Theory of Probabilities and Telephone Conversations. *Nyt Tidsskrift for Matematik*, B 20 (1909)
2. Erlang, A.K.: Solution of some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges. *Elektroteknikeren* 17 (1917)
3. Erlang, A.K.: Telephone Waiting Times. *Matematisk Tidsskrift*, B 31 (1920)
4. Cobham, A.: Priority Assignment in Waiting Line Problems. *Journal of the Operations Research Society of America* (1954)
5. Jackson, J.R.: Jobshop like Queueing Systems. *Management Science* 10 (1963)
6. Gordon, W.J., Newell, G.F.: Closed Queueing Systems with Exponential Servers. *Operations Research* 15(2) (1967)
7. Buzen, J.P.: Computational Algorithms for Closed Queueing Networks with Exponential Servers. *CACM* 16(9) (1973)
8. Baskett, F.T., Chandy, K.M., Muntz, R.R., Palacios, F.G.: Open, Closed and Mixed Networks with Different Classes of Customers. *JACM* 22(2) (1975)
9. Reiser, M., Kobayashi, H.: Queueing networks with multiple closed chains: Theory and computational algorithms. *IBM J. Res. Develop.* 19, 3 (1975)
10. Reiser, M., Lavenberg, S.S.: Mean Value Analysis of Closed Multichain Queueing Networks. *JACM* 27(2) (1980)
11. Denning, P.J., Buzen, J.P.: The Operational Analysis of Queueing Network Models. *ACM Computing Surveys* 10(3) (1978)
12. Buzen, J.P., Denning, P.J.: Operational Treatment of Queue Length Distributions and Mean Value Analysis. *Computer Performance* 1(1) (1980)
13. Courtois, P.J.: Decomposability: Queueing and Computer Systems Applications. Academic Press (1977)
14. Marie, R.: Modélisation par Réseaux de Files d'Attente. Thèse de Docteur ès Sciences Mathématiques. Université de Rennes (1977)
15. Gelenbe, E., Mitrani, I.: Analysis and Synthesis of Computer Systems. Academic Press (1981)
16. LINPACK: <http://www.netlib.org/linpack/>
17. SPEC: Standard Performance Evaluation Corporation, <http://www.spec.org/>
18. OPNET: Applications and Network Performance, <http://www.opnet.com/>
19. ns-2: The Network Simulator, <http://www.isi.edu/nsnam/ns/>
20. Barroso, L.A., Hözle: The case for energy proportional computing. *IEEE Computer* 40, 12 (2007)
21. COST IC0804. Energy efficiency in large scale distributed systems, <http://www.cost804.org/>
22. Chen, G., He, W., Liu, J., Nath, S., Rigas, L., Xiao, L., Zhao, F.: Energy-aware server provisioning and load dispatching for connection-intensive internet services. In: NSDI, San Francisco, CA (2008)
23. Standard Performance Evaluation Corporation (SPEC). SPEC power, <http://www.spec.org/>

24. Fan, X., Weber, W.D., Barroso, L.A.: Power provisioning for a warehouse-sized computer. In: Proceedings of the International Symposium on Computer Architecture, ISCA (2007)
25. Heath, T., Diniz, B., et al.: Energy conservation in heterogeneous server clusters. In: Proceedings of the Symposium on Principles and Practice of Parallel Programming, PPOPP (2005)
26. Chun, B.G., Iannaccone, G., Katz, R.H., Lee, G., Niccolini, L.: An Energy Case for Hybrid Datacenters. In: Workshop on Power Aware Computing and Systems, HotPower 2009 (2009)
27. Flinn, J., Satyanarayanan, M.: Energy-aware adaptation for mobile applications. In: SOSP 1999: Proceedings of the 17th ACM Symposium on Operating Systems Principles (1999)
28. Mankoff, J., Kravets, R., Blevis, E.: Some computer science issues in creating a sustainable world. *IEEE Computer* 41(8) (2008)
29. Im, C., Ha, S.: Energy optimization for latency- and quality-constrained video applications. *IEEE Design and Test of Computers* 21(5) (2004)
30. Iyer, S., Luo, L., Mayo, R., Ranganathan, P.: Energy-adaptive display system designs for future mobile environments. In: ACM MobiSys (2003)
31. Gilly, K., Alcaraz, S., Juiz, C., Puigjaner, R.: Analysis of burstiness monitoring and detection in an adaptive Web system. *Computer Networks* 53(5) (2009)
32. Executive Summary. Chapter 9: Projections of Future Climate Change. *Climate Change 2001: The Scientific Basis* (2005), http://www.grida.no/climate/ipcc_tar/wg1/339.htm
33. Feng, X., Ge, R., Cameron, K.W.: Power and Energy Profiling of Scientific Applications on Distributed Systems. In: Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium, IPDPS 2005 (2005)
34. Kumar, R., Farkas, K.I., Jouppi, N.P., Ranganathan, P., Tullsen, D.M.: Single-ISA heterogeneous multi-core architectures: The potential for processor power reduction. In: ACM/IEEE MICRO (2003)
35. Pinheiro, E., Bianchini, R., et al.: Load balancing and unbalancing for power and performance in cluster-based systems. In: Proceedings of the Workshop on Compilers and Operating Systems for Low Power, COLP (2001)
36. Ranganathan, P., Rivoire, S., Moore, J.: Power modeling and measurement. In: *Advances in Computers*. Elsevier (2009)
37. U.S. EPA. Report to congress on server and data center energy efficiency. Tech. Rep. (2007)
38. Kusic, D., Kephart, J.O., Hanson, J.E., Kandasamy, N., Jiang, G.: Power and performance management of virtualized computing environments via lookahead control. *Cluster Computing* 12(1) (2009)
39. Alcaraz, S., Gilly, K., Juiz, C., Puigjaner, R.: Handling HTTP flows over a DiffServ framework. In: LANC 2007 (2007)
40. Narayanan, D., Donnelly, A., Rowstron, A.: Write off-loading: Practical power management for enterprise storage. In: FAST (2008)
41. Nedeveschi, S., Popa, L., Iannaccone, G., Ratnasamy, S., Wetherall, D.: Reducing network energy consumption via sleeping and rate-adaptation. In: NSDI (2008)
42. Park, S., Jiang, W., Zhou, Y., Adve, S.: Managing energy-performance tradeoffs for multithreaded applications on multiprocessor architectures. In: SIGMETRICS (2007)
43. Raghavendra, R., Ranganathan, P., Talwar, V., Wang, Z., Zhu, X.: No power struggles: A unified multi-level power management architecture for the data center. In: ASPLOS (2008)

44. Menasce, D.A., Almeida, V., Dowdy, L.: Performance by Design: Computer Capacity Planning by Example. Prentice Hall PTR (2004)
45. Standard Performance Evaluation Corporation (SPEC). SPECweb2009, <http://www.spec.org/web2009/>
46. Menasce, D.A., Almeida, V.: Capacity Planning for Web Services: metrics, models, and methods. Prentice Hall PTR (2001)
47. Killelea, P.: Web Performance Tuning. O'Reilly (2002)
48. National greenhouse gas inventory data for the period 1990-2007. UN FCCC (2009), <http://unfccc.int/resource/docs/2009/sbi/eng/12.pdf>
49. Xie, F., Martonosi, M., Malik, S.: Compile-time dynamic voltage scaling settings: opportunities and limits. SIGPLAN Not. 38(5) (2003)
50. Banerjee, P., Patel, C.D., Bash, C., Ranganathan, P.: Sustainable Data Centers: Enabled by Supply and Demand Side Management. In: DAC 2009. ACM Press (2009)
51. Katz, R.: Tech Titans Building Boom. IEEE Spectrum (February 2009)
52. Patel, C.D., Ranganathan, P.: Enterprise power and cooling. ASPLOS Tutorial (2006)
53. Laudon, J.: Performance/Watt: The new server focus. SIGARCH Computer Architecture News 33(4) (2005)
54. Sun Microsystems. SWaP (Space, Watts and Performance) metric, <http://search.sun.com/main/>
55. Rivoire, S., Shah, M.A., Ranganathan, P., Kozyrakis, C.: JouleSort: A Balanced Energy-Efficiency Benchmark, <http://www.hpl.hp.com/environment/datacenters.html>
56. Molero, X., Juiz, C., Rodeño, M.J.: Evaluación y Modelado del Rendimiento de Sistemas Informáticos. Pearson Prentice Hall (2004)
57. Deloitte España. Barómetro de Empresas (2009).
58. Pierson, J.M., Hlavacs, H. (eds.): Proceedings of the COST Action IC0804 on Large Scale Distributed Systems, 1st Year (2010)
59. Slack, N., Chambers, S., Betts, A., Johnston, R.: Operations and Process Management: Principles and practice for strategic impact. FT Prentice Hall (2005)