Cumhuriyet Üniversitesi Fen Fakültesi
Fen Bilimleri Dergisi (CFD), Cilt:36, No: 6 Özel Sayı (2015)
ISSN: 1300-1949

Cumhuriyet University Faculty of Science
Science Journal (CSJ), Vol. 36, No: 6 Special Issue (2015)
ISSN: 1300-1949

ENERGY EFFICIANCY METRICS AND TECHNIQUES IN CLOUD DATA CENTERS

Shima Sokout JAHROMI[1,*], Mansourr AMINILARI[2], Amin TOUSI[2]

[1]*Department of Computer Engineering, Fars Science and Research Branch, Islamic Azad University, Marvdasht, Iran*

[2]*Department of Computer Engineering, Marvdasht Branch, Islamic Azad University, Marvdasht, Iran*

Received: 24.04.2015; Accepted: 09.07.2015

Abstract. Data centers are one of the major components of ICT field that have had significant growth to meet the needs of this area. In addition, the data centers are a computational resource for cloud computing and to have a good response time for a large number of their customers are often comprised of thousands of servers. In such a large scale, the energy consumption of data centers has increased extraordinarily. Increasing the power of consumption leads to increasing operational costs as well as greenhouse gas emissions. Thus, optimizing energy consumption in data centers is essential to reduce operational costs and protect the environment. Servers consume considerable amount of energy in cloud data center, thus optimizing the energy consumption of them has a significant role in reducing energy consumption. This paper provides a comprehensive study of the green metrics in the fields of energy efficiency in data centers, then classifies energy optimization approaches of servers in the data center and also provides a comprehensive review of techniques and their applications in each approach of recent research.

Keywords: Cloud computing, data center, energy efficiency, performance metric

INTRODUCTION

The primary focus of the designers of computer systems is to improve the efficiency and performance of the systems. According to this goal, the system performance due to the design of more efficient systems and by increasing the density of components is described by Moore's Law is steadily growing. Although the performance per watt ratio is increasing, the total energy consumed by computer systems is hardly reduced. On the opposite Total energy consumption is increased per year [1]. If this trend continues, the energy cost of the server during its lifetime will exceed the cost of hardware [2]. The problems for large-scale computing infrastructures, such as clusters and data centers will be worse. Data centers to meet the needs of a great number of customers often comprise of thousands of servers. Regarding the emergence of cloud computing and big data has increased the number of servers in data centers. At such a scale, the energy consumption of data centers and the subsequent operating costs will increase dramatically [3].

Major cloud providers like Microsoft, Amazon and Google have the largest and best equipped Cloud Data Centers. Each data center contains hundreds of thousands of servers, cooling equipment and transformer power amplifiers. For example, Microsoft data center in Quincy, Washington has 43,600 square meters of space, 4.8 kilometers of chiller piping, 965 kilometers of electric wire, 92,900 square meters of drywall and 1.5 tons of backup batteries. Although Microsoft as well as other companies did not release the number of servers on this site, but it announces that the data center consumes 48 megawatts which is equal to power 40,000 homes [4]. For another example, we can refer to energy consumption and carbon dioxide emissions of Google search engine. To respond to any search, Google uses its data centers containing thousands of servers. These servers consume a lot of energy

* Corresponding author. *E-mail:* shima.sokut@yahoo.com

Special Issue: International Conference on Non-Linear System & Optimization in Computer & Electrical Engineering

to perform the desired actions. "Figure 1" shows energy consumption and carbon dioxide emissions for

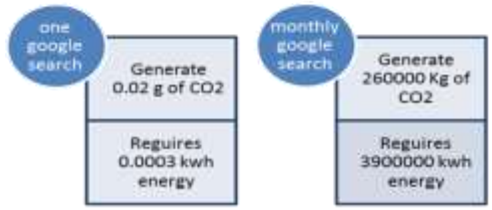

Figure 1. Google energy consumption.

A response to one Google search and monthly Google search. Google's monthly energy consumption is the equivalent to the average power consumption of 4239 homes for a month and the amount of carbon dioxide is equivalent to a car in 1,008,264 kilometers [5].

Rising the energy prices and extraordinary growth in data centers, strategies for reducing energy consumption have become more important. Problems such as rising energy costs, regulatory requirements and environmental and social issues about greenhouse gas emissions have increased the necessity and importance of further reducing energy consumption in data centers. Moreover, when the size of data centers increase, electricity costs cover other costs of cloud in all aspects. These indicate required research on methods for reducing energy consumption in data centers as well. In this paper, we will study metrics of energy efficiency, then review and compare existing approaches of optimizing the energy consumption of servers, techniques and challenges proposed in every approach, and applications in recent research.

Energy in data centers

The first step to saving energy of data centers is to gain a solid understanding of data center energy consumption. "Figure 2" shows the rate of energy consumption in a data center. As shown in "Figure 2", 30% of the input power consume in IT equipment.

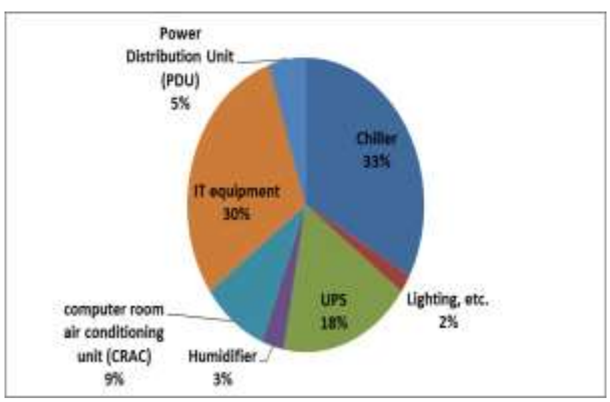

Figure 2. Distribution of energy consumption in data centers [6].

"Figure 3" shows the energy consumption of IT equipment. As you can see the servers consume a large amount of energy. So the use of techniques to reduce energy in servers will have a significant effect on energy savings.

Energy Efficiency Metrics

A. Power Usage Effectiveness (PUE) is the most common metric used to determine the energy efficiency of a data center [7]. PUE defined as follows.

$$PUE = \frac{\text{Total Facility Power}}{IT\ Equipment\ Power} \quad (1)$$

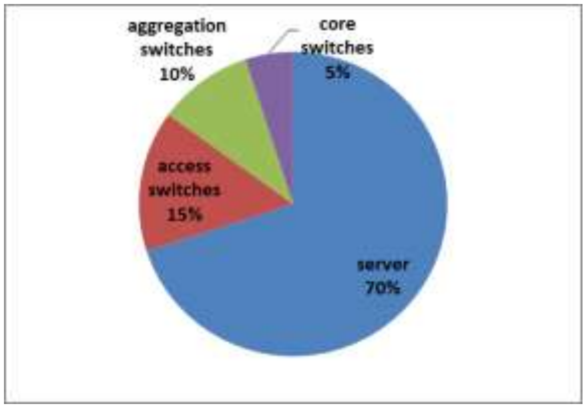

Figure 3. Distribution of energy consumption in IT equipment in data centers [8].

PUE represents the amount of energy used for IT infrastructure, cooling it and distribution systems required to maintain the availability and reliability of data centers. PUE is a very complicated metric. To understand of the metric, understanding load components and classification of measurement is very important [9]. In the following, components and sub-divisions of this metric as well as different levels and categories described in detail:

IT equipment power requirements: This energy includes the energy consumption of IT devices such as computing, storage device and networking with their complementary equipment such as switches, monitors, workstations and laptops will be used for monitoring or control purposes.

Total Energy Facility: This energy includes all energy consumed by IT equipment as described above, plus every energy used to support IT equipment such as:

- Power supply components: uninterrupted power supply, generators, power distribution units and batteries.
- Cooling systems Components: chillers, computer room air conditioning units (CRACs), direct expansion air handler (DX) units, pumps and cooling towers.
- Other energy consumption in data centers: server room light.

PUE is no upper limit and in practice, the facility load is always a little more than the IT equipment load. Smaller PUE shows more accurately by selecting a facility [10]. In a study carried out on 108 data center PUE have been reported 1/91 [11]. In a survey conducted on data center with 1100 end users around the world PUE have been reported between 1.8 and 1.9 [12].

B. Data Center Infrastructure Efficiency (DCiE) [13] is another common metric to measure the efficiency of energy consumption in data centers. Its computation formula is shown in Equation 2 and 3:

$$\text{DCiE} = \frac{1}{\text{PUE}} \quad (2)$$

$$\text{DCiE} = \frac{\text{Total Facility Power}}{\text{IT Equipment Power}} \quad (3)$$

C. Data Center Productivity (DCP) [13] is a quantity that shows the amount of useful work in a data center and calculated according to Equation 4. Where Tresource represents total resource taken to produce this useful work

$$\text{DCP} = \frac{\text{Useful work done}}{\text{T}_{\text{resource}}} \quad (4)$$

It is clear from the above metric that this metric defines a family of metric, because total resource can be any hardware or software resource. This family of metrics called DCxP. One of the most famous metrics in this family is Data Center Energy Productivity (DCeP) that shows the amount of useful work in terms of energy consumption of the data center. This metric calculated according to Equation 5:

$$\text{DCeP} = \frac{\text{Useful work done}}{\text{T}_{\text{Energy}}} \quad (5)$$

D. Compute Power Efficiency (CPE) [14]: This metric shows the overall efficiency of a data center. As shown in following Equation.

$$CPE = \frac{\text{IT Equipment Utilization}}{\text{PUE}} \quad (6)$$

$$CPE = \frac{\text{ITEquipment Utilization} * \text{IT Equipment Power}}{\text{Total Facility Power}} \quad (7)$$

The maximum value of this metric is one or one hundred percent. This metric has no unit so it provides comparing the overall efficiency of data centers.

E. HVAC stands for Heating, Ventilation, and Air Conditioning. The metric shows the rate of IT equipment power to the HVAC system energy. The HVAC system energy is the sum of the electrical energy for cooling, fan movement, and any other HVAC energy use like steam or chilled water. The HVAC metric is calculated as shown in Equation 8, where IT represents the annual IT electrical energy use and HVAC represents Annual HVAC electrical energy use. Fuel represents an annual fuel energy use. Steam represents annual district steam energy use and finally Chilled Water represents an annual district chilled water energy use [15]. It can be computed through equation 8.

$$\text{HVAC} = \frac{\text{IT}}{(\text{HVAC}+(\text{Fuel}+\text{Steam}+\text{ Chilled Water})\times 293)} \quad (8)$$

F. AxPUE [16] define a family of application level PUEs. The family's most famous metrics are Application Performance Power Usage Effectiveness (ApPUE) and Application Overall Power Usage Effectiveness (AoPUE).
ApPUE shows the energy efficiency of IT devices. This metric shows the rate of productivity applications to the IT devices from the start of the application to the

Completion and calculated according to Equation 9:

$$\text{ApPUE} = \frac{\text{Application performance}}{\text{IT Equipment Power}} \quad (9)$$

AoPUE metric shows data center energy consumption to increase productivity and calculated according to Equation 10:

$$\text{AoPUE} = \frac{\text{ApPUE}}{\text{PUE}} \qquad (10)$$

G. Data Center Workload Power Efficiency (DWPE) [17] shows the energy efficiency of High-Performance Computing (HPC) data center for a particular workload. This metric calculated according to Equation 11. Where Workload Power Efficiency (WPE) is a Performance/W metric for a HPC system and calculated according to Equation 12. System PUE (sPUE) shows the energy efficiency of a specific HPC system for a specific data center.

$$\text{DWPE} = \frac{\text{WPE}}{\text{sPUE}} \qquad (11)$$

$$\text{WPE} = \frac{\text{average achieved performance}}{\text{average HPC system power used}} \qquad (12)$$

Approaches to reduce energy consumption in servers

A. Power-proportionality: The main objective of this approach is to minimize the power consumption on servers to provide services to their customers' requests [18]. According to this approach, when the server load is equal to X% of maximum load, energy consumption should be X% of peak power [19]. This idea has led to the emergence of several techniques that follows:

1) Race to Idle: In this method the workload responses continuously at a maximum speed and then the server is idle until a certain amount of workload comes to the server again and the process is repeated [20].

2) Dynamic performance scaling (DPS) [21]: It adjusts performance components proportional to the energy efficiency dynamically. Dynamic voltage and frequency scaling (DVFS) and Dynamic Component Deactivation (DCD) are examples of this technique.

DVFS is an important tool in managing the balance between power consumption and performance of the CPU and reduces energy consumption of server from 10 to 50 percent [22].

DCD uses two approaches of probability and prediction to disable the components. In the prediction approach uses the past behavior of the system for future and in the probability approach models the behavior of the system as a possible model like Markov chain model [21].

B. Designing of energy-efficient servers: Most servers are designed to provide the highest levels of reliability, performance and scalability. This flexibility causes the servers won't be energy efficient. The primary objective of this approach is to design an efficient data center in terms of energy consumption. This approach will include the proposed techniques:

1) Using servers that are well designed to suit the particular application. In addition, these servers can be designed to be energy aware. For example, the design of the servers that consume less power in idle mode [23].
2) Using two running modes. One case with high efficiency and high energy consumption and the other case with low efficiency and low power consumption that server goes to one of the mode based on workload requirements [24].

3) Using a cluster of servers. This technique offers use of multiple servers with lower power and energy consumption instead of using one server with high power and a lot of energy consumption [25]. Of course, the software running on multiple servers, involved the challenge of distributing a larger server in several smaller servers. Moreover, more parallelism makes scalability difficult [26].

C. Providing server dynamically: The basic idea of this method is proportionally the number of servers with the workload. This approach aims to minimize energy consumption by powering the minimum number of servers required to meet the needs of a given workload. The advantage of this approach is that the server in off or sleep mode consumes less energy than idle state. The major disadvantage of this method is that it requires a few minutes to reboot the server and consumes a lot of energy [27]. This approach will include the proposed techniques:

1) Workload consolidation: This approach consolidates Workload of application on fewer numbers of servers. This consolidation optimizes energy consumption and increases efficiency.
2) Dynamic Power Management (DPM) [21]: It turns off Components dynamically based on the current workload.

D. Consolidation and virtualization: Server consolidation based on virtualization reduces energy consumption. This approach makes it easy to manage and reduce the cost of physical infrastructure [28]. The basic idea is, integrating multi-server workload on a server and uses a smaller number of servers. This approach allows turning off unnecessary servers and thereby energy consumption reduces and system efficiency increases [29].

The major difference between traditional data centers and cloud computing data centers is using virtualization. Virtualization is a fundamental element of cloud computing model that hides the complexity of this computing model from the perspective of users. With this technology the physical machines convert to several virtual machines and each virtual machine will meet the needs of several users and customers. Thus, virtual machines are the smallest shared part of the developer and the service provider so They have become an abstract level to provide the programs. It also provides access to shared resources, thus enabling optimal use of them [30].

Analysis of approaches energy efficiency

The power-proportionality approach is proposed first. Its purpose is to reduce unnecessary power consumption in the data center servers. But it should be considered that achieving the appropriateness of energy consumption with workload in hardware is very difficult.

The next approach is the design of energy efficient servers. Due to the dynamics of the workload, designing and customization of Server is very difficult and impact on data center Performance.

The third idea is to use the server dynamically. Have a good pattern of the input workload and convert it to a number of servers are the challenges of this approach. The boot time also has a negative impact on performance.

The last approach is consolidation and virtualization that are widely used in cloud computing servers. This approach requires the identification of patterns of applications use of sources that is difficult due to the dynamic nature of the workload. This approach uses the shared space so providing security is very difficult.

"Table 1" compares some Techniques to reducing energy consumption in cloud data centers In terms of energy savings, the impact on performance and response time.

Table 1. Analyze techniques of reducing energy consumption in cloud data centers.

Technique name	Energy Saving	Impact on Performance	Impact on response time
DVFS	Less than DPM	High for CPU bound program	High for CPU bound program
Virtualization	High	Low	Low
DPM	High	Low	Low
Workload consolidation	Depending on the size of the workload.	Depending on the size of the workload..	Depending on the size of the workload..

CONCLUSIONS

One of the key requirements of designing of modern computer systems such as data centers and cloud computing is energy efficiency, that is discussed due to the high energy consumption of IT field. Apart from cost, the amount of energy consumed will lead to a considerable expansion of carbon dioxide. Thus, optimizing energy consumption in cloud data centers is a very important issue in terms of environmental aspects and operational cost.

The energy metrics play a significant role for energy efficient performance of data centers. In this paper we have discussed various energy metrics to measure the energy efficiency of cloud data centers. Then, based on the fact that 30 percent of data center energy consumption related to IT equipment and of those, 70% of the energy used in servers, we discussed strategies for reduction of energy consumption in servers. We classified these strategies into four categories (power-proportionality, Designing of energy-efficient servers, Providing Server Dynamically and Consolidation and virtualization), then perused the existing techniques in both approaches completely. Finally, we analyzed and compared different aspects of the approach. Describing and comparing these approaches help people who wish to develop and improve these approaches.

REFERENCES

[1] Koomey, J.G. (2007). “Estimating Total Power Consumption by Servers in the US and the World”. Final report. 2 (3): 1-31.

[2] Barroso, L. (2005). “The Price of Performance”, ACM Press, 3(7): 48- 53.

[3] Sokout jahromi, S., Aminilari, M., Toosi, A. (2014). “Approaches to energy consumption management in cloud computing data centers", 4th International Conference on Information Technology Management, Communication and Computer, Tehran, IRAN, pp-183-191.

[4] Katz, R. H. (2009). “Tech Titans Building Boom.”. IEEE Spectrum, 46(2): 40–54.

[5] Wadhwa, B., and Verma, A. (2014). “Energy Saving Approaches for Green Cloud Computing: A Review”, IEEE Recent Advances in Engineering and Computational Sciences, 2(3): 1-6.

[6] Holzle, U., Clidaras, J., and Barroso, L. A.(2013). The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines (2nd Ed.). New York: Morgan and Claypool Publishers.

[7] Malone, C. Belady, C. (2006). “Metrics to Characterize Data Center & IT Equipment Energy Use”. Proceedings of 2006 Digital Power Forum, pp:34-45.

[8] Kulseitova, A., and Fong , T. (2013). “A survey of energy-efficient techniques in cloud data centers”. IEEE International Conference on ICT for Smart Society, Jakarta ,June, pp. 1-5.

[9] Dumitru. I., Fagarasan: I., Iliescu. S. Said, Y.H., and Ploix. S., 2011. “Increasing Energy Efficiency in Data Centers using Energy Management”, IEEE/ACM International Conference on Green Computing and Communications, Aug, pp. 159-165

[10] Azevedo, D., Cooley, J., Patterson, M. and Blackburn, M., 2011. “Data Center Efficiency Metrics: mPUE, Partial PUE, ERE, DCcE”. The Green Grid Presentation, 2(3): 1-37

[11] Sullivan. A. (2010). ENERGY STAR R for Data Centers. U.S. Environmental Protection Agency, presentation, on the WWW, URL http://www.energystar.gov/ia/partners/prod development/downloads/DataCenters GreenGrid02042010.pdf
[12] Stansberry, M., and Kudritzki, J.(2012). “Uptime Institute 2012 Data Center Industry Survey”. Technical report, See also URL http://uptimeinstitute.com/images/stories/Uptime_Institute_2012_Data_Industry_Survey.pdf
[13] The Green Grid in its white paper. (2007). “The Green Grid Data Center Power Efficiency Metrics: PUE and DCiE”. pp. 1-16. See also URL http://www.premiersolutionsco.com/wp-content/uploads/TGG_Data_Center_Power_Efficiency_Metrics_PUE_and_DCiE.pdf
[14] Belady, C.L., and Malone, CG. (2007).“Metrics and an Infrastructure Model to Evaluate Data Center Efficiency” ASME 2007 InterPACK Conference, 1, Vancouver, British Columbia, Canada, July, pp. 751-755.
[15] Mathew, P., Ganguly, S., Greenberg, S., and Sartor, D.(2009). “Self-benchmarking Guide for Data Centers: Metrics, Benchmarks, Actions”. Environmental Energy Technologies Division,2(4): 1-31.
[16] Zhou, R., Shi, Y., and Zhu, C.(2013). “AxPUE: Application Level Metrics for Power Usage”. IEEE International Conference on Big data , 5(4): 110-117.
[17] Wilde, T., Auweter, A., Pattersony. M.K., Shoukourian, H., Huber, H., Bodez, A., Labrenz, D., and Cavazzonix, C.(2014). “DWPE,a new data center energy-efficiency metric bridging the gap between infrastructure and workload”. IEEE International Conference on High Performance Computing & Simulation (HPCS), Bologna ,pp.893-901.
[18] Thereska, E., Donnelly, A., Narayanan, D.(2011). “Sierra: Practical Power-proportionality for Data Center Storage”. Proceedings of the sixth conference on Computer systems EuroSys '11, New York, NY, USA ,pp 169-182
[19] Barroso, LA., and Holzle, U.(2007). “The case for energy proportional computing”. ComputerJournal, 40(12): 33–37.
[20] Albers, S., Antoniadis, A.(2012). “Race to idle: new algorithms for speed scaling with a sleep state”. SODA '12 Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms, pp 1266-1285.
[21] Beloglazov, A., Buyya, R., Choon Lee, Y., and Zomaya, A.(2011). “A Taxonomy and Survey of Energy-Efficient Data Centers and Cloud Computing Systems”. Advances in computers, 8(82) 47-108.
[22] Rizvandi, N.B., Taheri, J., Zomaya, A.Y., and Lee, Y.(2010). “Linear Combinations of DVFS-Enabled Processor Frequencies to Modify the Energy-Aware Scheduling Algorithms. Cluster, Cloud and Grid Computing”. 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing (CCGrid) , Melbourne, Australia, pp. 388-397.
[23] Anagnostopoulou, V., Biswas, S., Saadeldeen, H., Savage, A., Bianchini, R., Yang, T., Franklin, D., and Chong. F.T.(2012). “Barely alive memory servers: Keeping data active in a low power state”, Journal on Emerging Technologies in Computing Systems, 8(4), pp 1-20.
[24] Gupta, V., Brett, P., Koufaty, D., Reddy, D., Hahn, S., Schwan, K. and Srinivasa, G.(2012). “HeteroMates: Providing high dynamic power range on client devices using heterogeneous core groups”. International Green Computing Conference, San Jose, CA, USA, pp 1-10.
[25] Vasudevan, V., Andersen, D., Kaminsky, M., Tan, L., Franklin,. J. and Moraru, I.(2010). “Energy efficient cluster computing with FAWN: workloads and implications”. 1st International Conference on Energy Efficient Computing and Networking, pages Passau, Germany, pp 195-204.

[26] Meisner, D., Wenisch, T.F.(2011). "Does Low-Power Design Imply Energy Efficiency for Data Centers?", International Symposium on Low Power Electronics and Design, Fukuoka, pp 109-114.

[27] Isci. C., McIntosh, S., Kephart, J., Das, R., Hanson, J., Piper,. S., Wolford, R,. Brey, T., Kanter, R., Ng, A., Norris, J., Traore, A., and Frissora, M.(2013). "Agile Efficient Virtualization Power Management with Low-latency Server Power States". ACM SIGARCH Computer Architecture News - ICSA '13, 41(3):96-107.

[28] Song, Y., Zhang. Y., Sun. Y., Shi. W.(2009). "Utility analysis for Internet-oriented server consolidation in VM-based data centers". IEEE International Conference on Cluster Computing and Workshops, New Orleans, LA, pp.1-10.

[29] Bari, Md. F., Boutaba, R., Esteves, R., Granville. L. Z., Podlesny. M., Rabbani. M. G., Zhang, Q., and Zhani, M. F.(2013). "Data Center Network Virtualization: A Survey". IEEE communications surveys & tutorials, 15(2): 909-928.

[30] Li, Y., Li, W., Jiang, C.(2010). "A Survey of Virtual Machine System: Current Technology and Future Trends". Third International Symposium on Electronic Commerce and Security, Guangzhou, pp. 332-336