Energy and Buildings 80 (2014) 562–569

Contents lists available at ScienceDirect

Energy and Buildings

journal homepage: www.elsevier.com/locate/enbuild

# Servers and data centers energy performance metrics

A.H. Beitelmal*, D. Fabris

*Department of Mechanical Engineering, Santa Clara University, Santa Clara, CA, USA*

ARTICLE INFO

*Article history:*
Received 13 April 2014
Accepted 16 April 2014
Available online 29 May 2014

*Keywords:*
Data center
Server efficiency
Energy
Cooling
Power
PUE
Efficiency metric
DC performance metric
DC productivity metric

ABSTRACT

A thermodynamic approach for evaluating energy performance (productivity) of information technology (IT) servers and data centers is presented. This approach is based on the first law efficiency to deliver energy performance metrics defined as the ratio of the useful work output (server utilization) to the total energy expanded to support the corresponding computational work. These energy performance metrics will facilitate proper energy evaluation and can be used as indicators to rank and classify IT systems and data centers regardless of their size, capacity or physical location. The current approach utilizes relevant and readily available information such as the total facility power, the servers' idle power, the average servers' utilization, the cooling power and the total IT equipment power. Experimental simulations and analysis are presented for a single and a dual-core IT server, and similar analysis is extended to a hypothetical data center. The current results show that the server energy efficiency increases with increasing CPU utilization and is higher for a multi-processor server than for a single-processor server. This is also true at the data center level however with a lower relative performance indicator value than for the server level.



## 1. Introduction

The continuous increase in data centers energy consumption and the inefficiency in data centers energy management are well-documented issues facing the Information and Communication Technologies (ICT) sector today and a key concern for Information Technology (IT) organizations worldwide. Industrialized countries use between 5 and 10% of their total electricity demand on their ICT sector [1] where data centers are the most important operation within ICT. Data centers account for about 25% of total corporate IT budget [1]. The total power consumption of data centers in the USA has been growing rapidly over the last 10 years. For example, in the Environmental Protection Agency (EPA) report to Congress the power consumed by the data centers in the USA was estimated at 61 billion kWh in 2006 [2]. This amounts to approximately 1.5% of the total electricity consumption in the USA in the same year [2].

It is not a surprise that the energy consumed to operate the cooling resources can easily exceed the energy consumed by the computational systems inside a data center. For example, about 45% of the total electrical energy used by a state-of-the-art data center is used for powering the cooling resources while about 40% is used to run the IT equipment and the rest or about 15% is used for power conversion & distribution [1]. Although over 30% of the servers housed within many data centers are practically idle (i.e. performing no useful computational work) however, they continue to consume energy to keep the servers uptime and availability as well as to maintain the proper thermal environment inside the data center. McKinsey and Company, an IT research and consultant company, suggests adopting new data center metrics to quantify energy efficiency in data centers [3]. The EPA moving in the same direction expressed the need to have the stake holders (the government, IT industry, data center operators and researchers) collaborate to achieve optimum energy efficiency in US data centers [2].

The demand for more and faster compute resources and higher uptime is increasingly forcing data center consultants and facility engineers to over-provision the cooling resources when designing new data centers. It is estimated that IT spending grows about 6% annually while the IT facility costs are growing by about 20%. Even during a rough economic climate, a significant percentage of corporate businesses are either planning or already building new data center facilities [3]. This indicates that the leadership of these businesses recognizes data centers as critical facilities to their present and future economic success. A typical data centers cost about $150 million to build in the early 2000s [3]. Now the same data center can cost several times as much with a substantial part of the cost related to the power and cooling infrastructure [3]. Data center reports show that the average daily servers'

* Corresponding author. Tel.: +1 4089055501.
*E-mail addresses:* abeitelmal@scu.edu (A.H. Beitelmal), dfabris@scu.edu (D. Fabris).

http://dx.doi.org/10.1016/j.enbuild.2014.04.036
0378-7788/

**Nomenclature**

| | |
|---|---|
| CRAH | computer room air handler unit |
| DC | data center |
| IT | information technology |
| $n$ | number of servers (systems) |
| $\bar{P}$ | average power consumption |
| $P_{chiller}$ | chiller power |
| $P_{comp}$ | compressor power |
| $P_{Cooling}$ | cooling resources power |
| $P_{CRAH}$ | CRAH power |
| $P_{CT}$ | cooling tower power |
| $P_{dyn}$ | dynamic power |
| $P_{IT,tot}$ | total IT power |
| $P_{Light}$ | data center lighting power |
| $P_{network}$ | networking systems power |
| $P_{management}$ | management-servers power |
| $P_{Pump}$ | chilled water pumping power |
| $P_{PDU}$ | power distribution unit power |
| $P_{Server}$ | server power |
| $P_{Storage}$ | storage racks power |
| $P_{Sup}$ | data center support (non-IT) power |
| $P_{tot}$ | total power |
| $P_{UPS}$ | uninterrupted power supply power |
| $\beta$ | server utilization |
| DCEPM | data center energy performance (productivity) metric |
| SEPM | server energy performance (productivity) metric |

utilization is less than 3% with maximum annual servers' utilization of 6% [3]. Overall a data center servers' utilization typically does not exceed 50%. Energy losses from systems on standby, over-provisioning of cooling resources and lack of robust systems and environmental monitoring and control can be as high as 50% of the total energy used by data centers [2–6]. The environmental impact in terms of $CO_2$ emissions and other harmful environmental pollutants produced by such power consumption is significant. To clarify this impact, a 50% loss of the EPA [2] estimated 120 billion kWh annual data centers energy consumption beyond 2011 by over-provisioning and inefficiency contributes an additional 52 million metric tons of $CO_2$ emissions based on an average of 0.87 kg $CO_2$ emitted per 1 kWh consumed. Recent studies estimate the carbon emissions from computing power (where data centers have the largest share) account for roughly 2% of the total global carbon emissions [1]. These indicators generated strong interest in developing intelligent energy consumption monitoring and measurement schemes to improve energy efficiency and generating energy savings that can translate into $CO_2$ emissions reduction and enhance the overall sustainability in this sector.

Data center cooling and power technologies have evolved slowly over the last 20 years, with only limited improvement to the old cooling and power technologies design approach. The low energy cost relative to the hardware cost is the main reason behind the cooling manufacturers' reluctance to alter the time-tested technologies, i.e., integration of new technologies on CRAH units, chiller systems and cooling towers. The changes in these systems were incremental, very slow and did not match the technological advancement in server design and computational applications. Although there is a strong desire on the part of the IT community to develop more "green" and sustainable approaches to data center design and computer operations, two main issues hinder progress: First, managing a large variety of systems and different levels of data center operations in a persistent and scalable manner is challenging; second, a lack of functional system and data center efficiency or energy performance metrics for operational energy consumption assessments and improvement.

A variety of published work on high density data centers power and thermal management issues are available. For example, Karlsson and Moshfegh [7] presented an investigation of the data center power requirement to maintain the proper indoor environment. They utilized an IR camera to visualize and show that the data center airflow lacks provisioning and will results in higher power consumption. Cho and Kim [8] numerically evaluated an aisle partition walls between cold and hot aisles within a data center. The results show that partition walls can improve performance efficiency based on general data center metrics selected for their study. Cho et al. [9] discussed major factors affecting airflow efficiency in data center and numerically evaluated airflow performance using multiple parameters and recommended further research on data center efficiency.

In the quest for improved efficiency, server manufacturers and other technology companies [10–17] have published technical reports or introduced new thermal management tools and technologies with the objective is to maximize data center energy efficiency and/or minimize the energy consumed by the cooling resources. Most of the current data center energy related solutions are custom-built for each data center and the authors are not aware of any published data that quantifies the actual data center benefits from these solutions. In addition, each of these data center solutions uses raw variables data such as temperature, relative humidity, pressure and power sensed from each system inside the data center. The large amount of data collected and the complicated algorithms needed to deal with this large amount of data could be overwhelming. These solutions would be easier to implement if system and data center performance and energy efficiency metrics were available as an intermediate step for savings quantification. For example these tools can use the data collected from the available sensors and calculate the energy performance metrics to determine their current productivity and energy efficiency level. This requires much simpler algorithms that uses the energy performance metrics rather than the raw data for evaluations, management and/or control. It is expected that the proposed metrics will complement current data centers' monitoring and management solutions and provide a way to evaluate and compare servers and data center performance. Factors such as system location, environmental conditions and other internal factors can potentially affect system performance. Previous work on data center efficiency did not take into account individual servers and rack performance in data center evaluation [15–17]. In addition, most of these evaluations can only qualitatively assess data center performance. Immediate attention to these gaps is necessary to quantify first the system level energy performance and second the data center level energy performance. The energy efficiency or performance will be defined through universal metrics that can quantify these terms and help rank and differentiate existing and future servers and data centers designs.

In 2011, Beitelmal and Fabris [18] introduced new energy efficiency metrics for servers and data centers. These metrics were based on the first law of thermodynamics thermal efficiency definition to facilitate proper evaluation of data centers power and energy efficiency; help reduce the total cost of operation and provision data centers energy resources. Their results show that these metrics are universally applicable and can be used to compare energy efficiency between different systems and data centers. In 2013, Yuventi and Mehdizadeh [19] published an article that supports Beitelmal and Fabris [18] approach on the energy efficiency metrics and their view on the alarming misuse of PUE as a data center metric. Furthermore, the article highlighted additional issues with PUE determinations.

Recently a number of United States senators introduced the Shaheen-Portman energy efficiency bill [20] to promote energy