

Energy Consumption Modeling and Prediction in the Cloud Data Centers

Sara Diouani* and Hicham Medromi

*Research Foundation for Development and Innovation in Science and Engineering, Casablanca, Morocco
Engineering research laboratory (LRI), System Architecture Team (EAS), National and High School of Electricity and Mechanic
(ENSEM), Hassan II University of Casablanca, Morocco*

Received 29 August 2019; Accepted 29 April 2020

Abstract

Energy efficiency has turned into an undeniably essential worry in data centers as a result of issues related to energy consumption, including capital costs, working costs, and natural effect. While energy wastage, because of imperfect utilization of facilities and IT hardware, has to a great extent been decreased using best-practice advances, tending to energy loss in IT equipment, still requires the plan and usage of energy-aware resource management frameworks. Various methods aimed to enhance the energy efficiency in data centers have been developed in the field related to the development of resource allocation. In fact, researchers and scientists utilize numerous ways to improve effectiveness for optimized power and performance: scaling virtual machine and server preparing abilities to minimize energy consumption, enhancing resource use through exploiting its heterogeneity and consolidating the workload. To accomplish these objectives, there is multiple proposes models, calculations, algorithms, techniques, and procedures that diminish energy use using virtual machine scaling, virtual machine sizing for CPU and memory, CPU frequency adjustment, and also equipment power control for resource allocation in the server-level. The proposed online performance and power models catch framework conduct while adjusting to changes in the fundamental infrastructure. In this context, we present different formulas for calculating energy consumption in cloud data centers and we propose a simplified and completed deduced formula for energy consumption estimation

Keywords: green cloud computing, data center, energy consumption, energy efficiency, resource allocation, optimization.

1. Introduction

The advancement of Internet technologies has offered to ascend to the cloud computing paradigm, a model that empowers adaptable, cost-effective, and scalable resource allocation while offering usability to end consumers. The expanding interest in cloud administrations has prompted quick development in the number of computing resources and the number of large-scale processing data centers.

Data centers can involve a huge number of server units that possess a lot of space and consume impressive measures of power. A single data center can occupy 100,000 m² space, contain 12,000 server racks, and host one server. As indicated by estimation by the Natural Resource Defense Council, 12,000,000 data center servers are assessed to convey about all the USA's online exercises in 2014 [1]

The data centers named "hyperscale" data centers will be the biggest data centers, and it will grow from 338 in 2016 to 628 by 2021, as appeared in figure 1, and which will represent approximatively by 2021 53% of all installed data center servers [2].

The energy consumption of data centers is likewise anticipated that would increment significantly: one estimate proposes the extent of power consumed by data centers all

over the world will increment from 1.3% in 2010 to 8% by 2020 [4]. The impacts of high-power utilization incorporate high energy costs for cloud suppliers, critical interest in the plant to cool servers in data centers, and high carbon dioxide emanations. Around the world, data centers acquire \$27 billion yearly expense of the energy required to control data centers [5]. Yearly, CO₂ outflows are anticipated to increment from 116 million metric tons in 2007 to 257 million metric tons in 2020 [6], and carbon emanations from data centers are relied upon to match or surpass that of the airline industry by 2020 [5].

The impacts of high power utilization incorporate high energy costs for cloud suppliers, critical interest in the plant to cool servers in data centers, and high carbon dioxide emanations. Around the world, data centers acquire \$27 billion yearly expense of the energy required to control data centers [5]. Yearly, CO₂ outflows are anticipated to increment from 116 million metric tons in 2007 to 257 million metric tons in 2020 [6], and carbon emanations from data centers are relied upon to match or surpass that of the airline industry by 2020 [5]. This enormous consumption of energy by data centers can be credited to some extent to the developing numbers of data centers and servers because of expanding interest for cloud administrations. But it additionally reflects wasteful utilization of the energy expended. Not all the energy that is conveyed to the data center is used to help the core activity of giving cloud computing services; a noteworthy part of the energy is devoured by cooling frameworks, un-

*E-mail address: Diouanisara19@gmail.com

ISSN: 1791-2377 © 2020 School of Science, IHU. All rights reserved.

doi:10.25103/jestr.133.25

interruptible power supplies, and so forth. In legacy data centers, up to half of the power is used by non-server hardware [7]. While these data centers presently can't seem to receive energy-efficient prescribed procedures in their structure and activity, in general, the industry has gained critical ground around there.

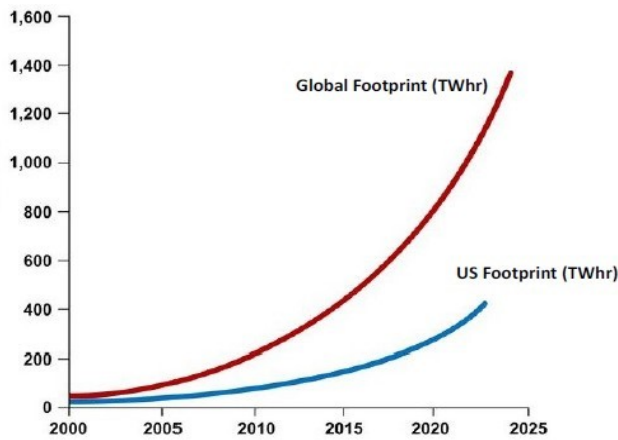


Fig. 1. Projection of electricity use in data centers [3]

In this way, chances to ameliorate the power proficiency of data centers presently to a great extent rely upon discovering approaches to enhance the energy efficiency of the servers in the data center grounded in the calculation of energy consumption. This paper is made out of four sections. Section II presents an overview of the background related to the data centers. Section III offers an outline of the context along with the studied problem and which are referred to data centers, performance, and issues associated with energy-efficiency in data centers. Section IV incorporate a review of related works in the areas of data center energy consumption calculation and the different formulas for calculating energy consumption in cloud data centers. In section V, a deduced simplified formula for calculating energy consumption in cloud data centers is proposed. Finally, the paper in the last part gives an outline of our commitments and their future aspiring expansions.

2. Background

Data centers commonly are powered by power electricity. Be that as it may, following the system for diminishing carbon emanations and conforming to manageable operational models, modern data centers utilize elective energy sources, for example, geothermal, wind and sunlight based power. The electric power streams from outside power grids into inner infrastructure facilities, Information Technology (IT) hardware and other support systems. The energy flows to the inner IT offices through Uninterrupted Power Supplies (UPS) to keep up a steady power distribution notwithstanding during conceivable power disappointments.

The design of a data center is complex and intricate since it does not just comprise of the equipment components yet, in addition, the software that keeps running in the IT framework infrastructure. Thusly, we can arrange its components into two layers which are hardware and software layers, as appeared in Fig. 3. The hardware or equipment comprises of multiple elements. The significant ones are cooling frameworks, power distribution units, lighting equipment, servers, and networking materials. The software layer can be additionally partitioned into two subcategories, the Operating

System/Virtualization layer, and the applications. The first refer principally to the host OS which is installed on the servers and the cloud deployment running over it. The second alludes to the distinctive sort of applications running in the servers which differ depending on the business and the industry cases.

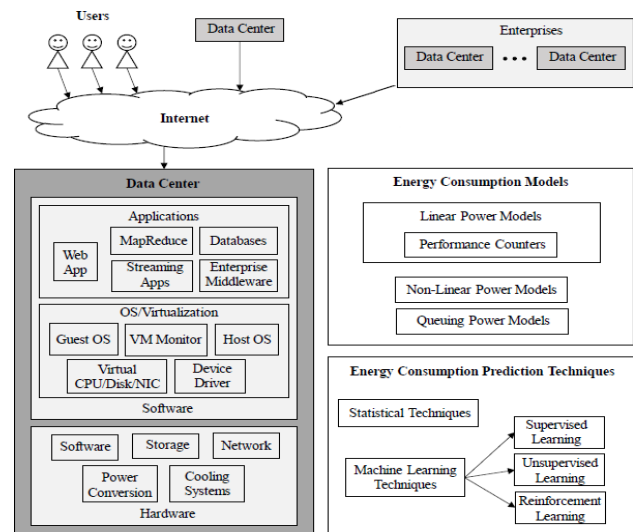


Fig. 2. A view with regards to energy consumption modeling and prediction in the cloud data centers layers (software and hardware layers).

Seeing how the energy is shared among the components of such a complex and an unpredictable system just as foreseeing energy utilization, requires a framework optimization cycle as exhibited in [11]. Regardless of whether we need to display the utilization of the entire data center or we are especially intrigued by the IT framework, the general methodology can be limited to the following procedure. At first, we have to measure the energy consumption of every part which is considered and distinguish where the most energy is expended. For that, we select the features that will develop the power model. Various procedures can be adopted for feature selection, for example, regression analysis. The precision of the power model should be approved. At long last, the model can be utilized to predict the energy consumption of the system and discover intends to concentrate on improving the energy efficiency of the data center.

There has been great deal of research on energy utilization prediction and expectation for Information Technology infrastructure of data centers. Activities and efforts to minimize the cost related with the power dissemination and cooling of the equipment are extending, hence the power management has turned into a fundamental issue in business environments. Server takes the most noteworthy part of the energy consumption in the racks [11]. Thus, there is need of considering how the energy is devoured and what are the principle parts of the server which effects the energy utilization the most. A few investigations [66] [67] assess power level consumption of the system and propose models which anticipate the energy considering different server components. Different studies [68] [69] [70] [71] center around processor's power use. The processor is the segment inside the server which expends the most energy.

Equipment Performance Counters are proposed in numerous works such as in [68] [69] [71] to predict power consumption of any processor with the utilization of data investigation or statistical methods. They give multiple information about the processor performance. The event

counter mechanisms have distinctive implementation which varies relying upon the processor family, and thus does the quantity of the accessible hardware and software events. Additionally, there are limitations on how numbers of events can be estimated all the while. For instance, the IBM Power 3-II has 238 accessible performance counters, while just 8 of them can be estimated at the same time [72] and, only 2 events out of 77 can be estimated simultaneously in the Intel Pentium II processor [72].

3. Context and Problem

Since data centers are tremendous scale and have critical computing infrastructures which are nonstop working [8], [9], they are pushing the quick development of the IT industry and change widely the economy.

The critical nature of data centers has been filled for the most part by two wonders. First, the regularly expanding development in the interest in data computing, processing, and storage capacity by an assortment of extensive scales cloud services, like Google and Facebook. Also, banks and some telecommunication providers like British Telecom [10], engender the multiplication of huge data centers with thousands or millions of servers.

Moreover, the necessity for withstanding a tremendous assortment of applications that keep running for a couple of moments and which run constantly on jointed equipment stages [8] has advanced building large scale computing infrastructures. Thus, data centers have been praised as one of the key empowering advances for the quickly developing the IT industry and in the meantime, bringing about the worldwide market size of 152 billion US dollars in 2016 [11]. Data centers being substantial-scale computing infrastructures have colossal energy budgets, that offered ascend to different problems related to energy efficiency.

In the first place, energy effectiveness of data centers has accomplished a major significance in latest years since it has a high monetary effect including (i) high financial, (ii) ecological, and (iii) performance effect. An ordinary data center can expend energy equal to 25,000 households.

The electricity consumption of the data center can exceed 100 times the equivalent of standard office space [5]. Moreover, the energy expenses power of an ordinary data center is duplicated every five years [8]. In this way, with such increment in power utilization and thus electricity expenditures increase, also, power payments have turned into a huge cost for the present data centers [12], [13]. Sometimes, these charges can surpass the expense of acquiring equipment [14].

Also, data center energy utilization makes various environmental issues [15], [16]. For instance, in 2005, the aggregate data center power utilization was 1% of the total power consumed in the United States and made as many discharges as a medium-sized country as Argentina [17].

In 2010 the worldwide power use by data centers was evaluated to be in the range of 1.1% and 1.5% of the aggregate overall power utilization [18], while in the United States about 1.7% to 2.2% of all US electrical use was devoured [19]. According to this report [20], in 2012, the data centers overall devoured about 270 TWh of energy and this utilization has a Compound Annual Growth Rate (CAGR) of 4.4%.

Thus, currently, data center energy efficiency is viewed as a major worry for data center administrators, in front of the conventional examinations including availability and security. At last, notwithstanding running servers in the idle

mode expend an important measure of energy. Extensive reserve funds can be done by turning off these servers. These different measurements, for example, workload consolidation should be taken to diminish data center power utilization. In the meantime, these power-saving methods diminish framework execution, indicating a harmony between energy reserve and high performance. Two sections of the energy devoured by a data center exist energy use by IT equipment (e.g., servers, networks, storage, and so forth.) and utilization by infrastructure facilities (including cooling and power conditioning frameworks).

The measure of energy consumed by these subcomponents rely upon the structure of the data center and also the efficiency of the equipment. In this paper, we cover a broad number of the different formula used in the modeling of energy consumption. A general way to deal with data center energy consumption comprises of four principal stages (see Figure 3): feature extraction, model development, model verification, and use of the model, for example, prediction.

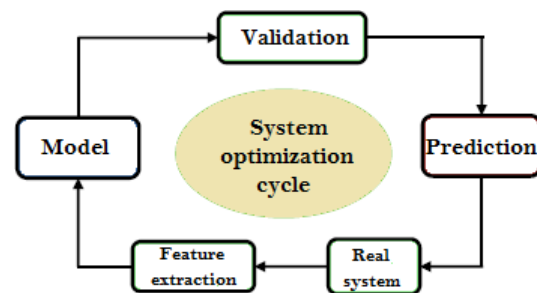


Fig. 3. The energy consumption modeling and the process of prediction [11]

In the feature extraction stage: To diminish the energy consumption of a data center, first, it is required to quantify the energy utilization of its segments [21] and recognize where the majority of the energy is wasted. This is the task of the feature extraction phase.

A model can be defined as a formal abstraction of a real system. Models for computer frameworks may be formulated as equations, graphical models, collection of representative examples, neural networks, and so forth. The decision of representation influences the exactness of the models, and their ability to interpret by individuals [22]. Exact power utilization models are extremely critical for multiple energy efficiency plans utilized in processing hardware [23]. There are various utilizations for power models:

- Design of data center frameworks: Power models are vital in the underlying plan of parts and frameworks because it is impracticable to assemble physical frameworks to evaluate each structure decision's impact on power utilization [23]. For instance, this methodology was adopted for the project of Data Center Efficiency Building Blocks [24].

- Forecasting the patterns in energy effectiveness: In every-day tasks of computer frameworks, clients and administrators of the data center have to comprehend the power use patterns of computer frameworks to augment their energy efficiency. Separate physical power estimation can not answer because it is not possible to anticipate future power utilization [25]. Estimated measurements additionally don't give a connection between resource utilization and power consumption [23]. Test confirmation using real experimental data is commonly costly and non-flexible. Energy models are also less expensive and increasingly adaptive to changes in operating parameters [26].

-Energy utilization improvement: Multiple distinctive power utilization enhancement plans have been created over power consumption models which are formulated as mathematical functions [27]. Power modeling is an active zone of research including examining relationships between the framework use and power utilization [28].

But, modeling the correct energy consumption conduct of a data center, at the entire framework level or the individual part level is quite difficult. Specifically, the data center utilization designs rely upon numerous components, for example, equipment details, workload, cooling prerequisites, types of applications, and so forth, which cannot be estimated with ease.

The power devoured by equipment, software that keeps running on equipment and the cooling and power framework of the building in which the data center frameworks reside are altogether firmly coupled [29]. Moreover, it is unrealistic to perform detailed estimations of the energy utilization of all lower level parts, since the estimation infrastructure introduces an overhead to the framework. Because of these reasons energy utilization expectation strategies have been created which can estimate the level of energy consumed by a system for a given workload. Energy consumption prediction methods can likewise be used for forecasting the energy use of a given data center working in an explicit setting.

Note that the context of our research will be limited by the study of energy consumption at the server level and we will neglect the power models which consider the System's heat generated by the server cooling and the server power dissipation.

When approved the model can be adopted for various tasks, for example, prediction of the energy consumption of the data center. The experiment picked up by the prediction of the energy consumption of a real framework can be used for enhancing the energy consumption model.

4. Related works: analysis and modeling of energy consumption in cloud data centers

Data centers are so critical since its energy-hungry infrastructures run large-scale Internet-based services. Energy consumption models are crucial in planning and optimizing energy-efficient operations to check over the excessive energy consumption in data centers. Nothing that, establishing an energy consumption model is related to establishing the relations between some parameters (such as CPU, memory, etc.) of resources and thus the energy consumption of the overall system. And in light of the targets of our research, we survey an in-depth study of the current research jointed to the areas of data center energy consumption and we present some different formulas for calculation, modeling, and prediction for data center energy consumption.

A. Energy vs Power

Energy (E) represents the aggregate sum of work achieved by a system in a period of time (T) and power (P) is the rate at which the work is executed by the system. The link between these three amounts is presented as follows:

$$E = PT \tag{1}$$

Where:

- E represents the energy consumption of the system measured in Joules, and P is estimated in Watts;
- T represents the period of time in seconds.

If T is quantified in-unit times (hour), so the estimations of energy (Watt-hour) and power (Watts) turn into equal. The following formulation can be marginally improved by considering energy as the integration of power values in a period beginning from $t1$ and closures at $t2$. It should be noted that we utilize the terms energy and power reciprocally in our work.

B. Server power models

In this subsection, we examine the additive power models which consider the whole power consumption of the server as a summation of its subcomponents.

One of the simpler power models was presented in [30], which designed the server power as a summation of the CPU and memory power utilization and this power model is as follows:

$$E(A) = E_{cpu}(A) + E_{memory}(A) \tag{2}$$

Where:

- $E_{cpu}(A)$: represents the energy consumption of the CPU;
- $E_{memory}(A)$: represents the energy consumption of the memory.

Other exhaustive power models have been made by taking into account other parts of a server, for example, disks, network peripherals, and so forth. Server energy utilization model depicted in [31] enlarges the following power model with Input/Output (I/O) parameters. This model can be appeared as:

$$E_{total} = E_{cpu} + E_{memory} + E_{I/O} \tag{3}$$

Where energy utilized by the server is formulated as a function of energy used by CPU, memory, and I/O devices. In any case, the vast majority of the present platforms do not permit estimating the power devoured by the three primary sub-frameworks (CPU, memory, and disk) of servers independently. Only the full system power presented by E_{total} may be estimated [32].

Other researchers have additionally presented an equivalent power model by modeling the framework power utilization as a summation of CPU, memory, and other framework components [33]. The power model depicted in equation (3) may be extended as follows [34]:

$$E_{total} = E_{cpu} + E_{memory} + E_{disk} + E_{NIC} \tag{4}$$

Where:

- E_{cpu} , E_{memory} , E_{disk} , and E_{NIC} : represent the energy expended by CPU, memory, disk, and network interface card (NIC) respectively.

Moreover, this model can consolidate an extra term for energy utilization of mother board as depicted in [35], [36] and [37] or a baseline constant, for example, presented in [38].

This energy model can be additionally extended by considering the way that energy can be estimated by duplicating average power with execution time as follows [34]:

$$E_{total} = P_{comp}T_{comp} + P_{NIC}T_{comm} + P_{net-dev}T_{net-dev} \quad (5)$$

Where:

- P_{comp} : signifies combined CPU and memory average power use;
- T_{comp} : is the average computation time;
- P_{NIC} : denotes the average network interface card power;
- T_{comm} : represents the total network time.

This energy model additionally considers the energy cost from network devices power $P_{net-dev}$ and the running time denoted by $P_{net-dev}$ when the devices are under load. Another form of this energy model can be developed by taking into account the levels of resource usage by the main parts of a server [39] as follows:

$$P_t = C_{cpu}u_{cpu,t} + C_{memory}u_{memory,t} + C_{disk}u_{disk,t} + C_{nic}u_{nic,t} \quad (6)$$

Where:

- U_{cpu} : refers to the CPU utilization;
- U_{memory} : denotes the memory access rate;
- U_{disk} : represent the hard disk I/O demand quota;
- U_{net} : is the network I/O demand amount.

Also, P_t is associated to the predicted power consumption of server at a time t , and the C_{cpu} , C_{memory} , C_{disks} , and C_{nic} refers to the coefficients of CPU, memory, disk and network interface card respectively.

An alternate kind of power models depend on the sort of tasks led by a server. In this methodology, which is like the power utilization of CMOS (Complementary Metal Oxide Semi-conductor) circuits, the energy consumption of data center systems is split into two segments called static power P_{fix} and dynamic power P_{var} [40], [41], [42], knowing that the division between static and dynamic power relies upon both the considered system and the workload itself. We can model it as follows:

$$P_{total} = P_{fix} + P_{var} \quad (7)$$

P_{fix} refers to is the static power utilization with regards to a server, and which is the power that is devoured by the system independent of its state of operation. This incorporates power lost due to discharge in semiconductor parts, for example, CPU, memory, I/O and more motherboard segments, and so on [43].

This classification additionally incorporates power required to keep fundamental working system processes and other idling tasks (e.g., power required to keep the hardware clocks, timer interrupts, disk drives active, and network ports [44]). The leaking currents should be kept least to maintain a strategic distance from such energy waste. But, this requires enhancement of the lower level (semi-conductor chip level) energy utilization [40].

Also, dynamic power utilization with regards to a server is built by movements, for example, the actions of circuits, connection to disc drives (I/O), and so forth. It relies on the most part on the category of workload which performs on the computer and in addition to how the workload adopts CPU,

memory, I/O, of the system and so on. [43]. Moreover, approximately about 30–40% of the power is consumed on the disk, the network, the I/O and peripherals, the regulators, etc.

In fact, the majority of the power models already depicted depended on segment wise power utilization decomposition. But, there can be other diverse sorts of energy utilization models created for a server dependent on its periods of execution. As an example, equation (8) is the energy model described by Orgerie et al. [45] (however, it is not specifically attributed to servers by them).

$$E = E_{boot} + E_{work} + E_{halt} \quad (8)$$

Where:

- E_{boot} : refers to system booting;
- E_{halt} : corresponds to halting energy utilization which is null if the equipment requires not be booting or halting during its activity life cycle.

But, the adoption of this sort of activity stage based energy models is quite uncommon in the real world. Instead, framework usage-based power models are intensely adopted in data center power modeling. The use of the virtual machine (VM) power as a parameter in the power model is adopted for modeling server power which is treated as another componentwise power breakdown approach. This power model of the server is presented as follows [46]:

$$P_{server} = P_{baseline} + \sum_{i=1}^n P_{vm}(i) \quad (9)$$

and:

- P_{server} : denotes the total power of the server;
- $P_{baseline}$: represents the baseline power which is empirically determined;
- P_{vm} : refers to the power of an active VM;
- n : represents the number of VMs hosted in the server.

And every single VM's power consumption can be modeled as:

$$P_{vm} = \alpha U_{cpu} + \beta U_{mem} + \gamma U_{io} + e \quad (10)$$

Where:

- U_{cpu} , U_{mem} , and U_{io} : denotes CPU utilization, memory usage, and disk IO throughput, respectively;
- e : represents an adjustment value.
- α , β , and γ : denotes the weights that should to be qualified off-line.

Indeed, this equation is relatively identical to the power model presented in equation (3), which likewise modeled the power of the server by the use of a component wise breakdown.

Thus, the total server power can be modeled as:

$$P_{server} = \alpha \sum_{k=1}^n U_{cpu}(k) + \beta \sum_{k=1}^n U_{mem}(k) + \gamma \sum_{k=1}^n U_{io}(k) + ne + E_{baseline} \quad (11)$$

and:

$-n$: represents the number of VMs in the physical host.

X. Xu et al. [47] conceived a similar power model for server by taking into account CPU, disk and idle power consumption. In this model, CPUs and disks are treated as the considerable parts which mirror the framework movements [47].

Also, Z. Zhang et al. [48] aimed to create a platform autonomous way to deal with the system for power consumption estimation dependent on hardware achievement measurements in virtualized cloud computing, which can give evaluations to control the power management.

Basically, because of the isolation, it is not possible to estimate the performance measurements of explicit equipment from the host operating system in the time when a part of escalated workloads are running inside VM instances. So performance metrics in VM level are gathered and compiled to evaluate power consumption for an individual server.

In a general sense, the power use can be found in two viewpoints. The first is the static power utilization which is principally founded on the hardware configuration and mirrors the idle power utilization of a framework. For the most part, static power utilization is easy to model for disposed equipment setup, which around record for close to half of maximum power utilization. The other detail is the dynamic power utilization produced by running extra intensive workloads. These additional workloads may powerfully change uses of power-aware equipment devices.

According to [48], CPU, memory and hard disk are considered as the ultimate three powerful and dynamic power expending devices in the computer. Regarding CPU, there are many variables identified with power utilization.

In the system proposed in [48], CPU usage and CPU frequency are selected as basic power-aware elements. In virtualized computing framework, CPU resources are linked to VM instances according to the software configuration of VM hypervisor. VM hypervisor is responsible for resource allocation among VM instances.

CPU frequency may affect the power use of maximum devices in a computer and high CPU frequency will influence RAM and hard disk to exhaust more power. Memory is another noteworthy power-aware device. Yet, memory use isn't totally immediate compared with power use of memory and this problem is tended to by viewing the memory utilization as a distinct rate rather than a persistent one. I/O use is indicated by bytes written and read every second, which may, generally, emulate the power utilization in I/O channel. In light of these I/O performance measurements, the calculation of the power consumed by hard drive may be estimated.

By the use of the performance measurements, the model for server power calculation might be composed as [48]:

$$P_w(k) = w_{cpu}U_{cpu}(k) + w_{mem}U_{mem}(k) + w_{io}U_{io}(k) + \varepsilon \quad (12)$$

Where:

- w_{cpu}, w_{mem}, w_{io} : represent the linear coefficients;
- ε : denote the error term of linear model;
- $U_{cpu}(k), U_{mem}(k), U_{io}(k)$: represent the overall use of CPU, memory, and I/O in the time stamp k .

In this design [48], the power utilization of other different resources is not taken into consideration, on the grounds that

their dynamic scope of power consumption is generally limited and can be incorporated into the error term.

The power utilization in an individual server is determined by including resource uses of all the residing VMs collectively and implementing these data to determine power calculation as:

$$P_w(k) = \sum_{j=1}^N P_{w_j}(k) \quad (13)$$

Where:

- $P_{w_j}(k)$: signifies the power use of a single server;
- N : indicates the number of physical servers in the framework.

Likewise, the performance can be introduced as an overall performance by discovering some parameters, for example, an application running around or overhead in an individual VM and include them together as:

$$P_f(k) = \sum_{i=1}^N P_{f_i}(k) \quad (14)$$

Where:

n : illustrates the number of VM instances in the system.

The VM Consolidation (VMC) is a procedure fused in Cloud Resource Management System (CRMS) to build the energy proficiency of Cloud. Equipment failure of existing PMs and expansion of new ones are constant occasions in the data center. Besides, a resource necessary to achieve the rest of the assignments of existing service demands is in a continuous increase. Subsequently, as time advances, remapping of residual workload to currently available resources wind up unavoidable to maintain the enhancement of Cloud resource utilization. The VMC method is adapted to move available VMs into a minimum number of dynamic physical machines (PMs), so the PMs with no VM can be maintained into a sleep state. And since energy consumption by the PM in a sleep state is fundamentally reduced than the energy consumption by the PM in an active state. Thus, by VMC, the energy consumption of Cloud data center is reduced.

The more number of VMs are put at one PM, the lesser turns into the general energy consumption with a minimum number of active servers. Thus, resource use rate of the PM P_i, R_{P_i} (4) would become bigger which thus would increment the resource usage proportion of cloud data center (CDC) R_{CDC} (5).

$$R_{P_i} = \frac{\text{Used measure of resource of } P_i}{\text{Total measure resource of } P_i} \quad (15)$$

And:

N : indicates the total number of active hosts in CDC.

$$R_{CDC} = \frac{1}{N} \sum_{i=1}^N R_{P_i} \quad (16)$$

In [49], the proposed model for energy optimization in the cloud data center uses different controllers and allocation techniques. Also, to characterize the performance metrics needed to define the state of a PM, the efficiency of a disposed resource (P_i). The researchers considered the utilization proportion of the (CPU), the use proportion of the fundamental memory (MEM) and the energy consumption of a resource ($CONS_i$). Also, identified with VM, they considered the inner level of CPU and memory utilization, but not the rate dissipated by the VM on the PM. Additionally, according to [49], the estimation of the efficiency of every resource by considering the PMs can be defined.

$$E_{CPU} = \frac{CPU_{PM}}{CONS_{CPU_{PM}}} E_{MEM} = \frac{MEM_{PM}}{CONS_{MEM_{PM}}} \quad (17)$$

At the PM stage, every PM will host various VM, and each indicates a user. These VM don't use data in common, they are self-reliant. Every PM memorizes a record of states, where a state is characterized as a vector that encloses the above parameters:

Where:

- CPU_{PM} : indicates the CPU use of a PM;
- MEM_{PM} : indicates the memory use of a PM;
- $CONSCPUPM$: represents the power consumption of a PM in relation to CPU;
- $CONSMEMPM$: expresses the power consumption of a PM in relation to memory;

The Autonomic energy-aware tasks scheduling approach proposed by researchers in [50], the metric considered to be minimized for each host H_j by considering the CPU as the primal parameter, is represented as follows (18):

$$WPN = \frac{\sum_{i=1}^{NbCPU} ((P_{max}^{CPU_i} - P_{min}^{CPU_i}) \times \frac{\sum_{k=1}^{NbVM} M_{max}^{VMK}}{H_j})}{NbCPU \cdot M_{max}^{H_j}} \quad (18)$$

Also:

$$P_{min}^{cpu_i} = \frac{P_{min}^{host j} \times M_{curr}^{cpu_i}}{H_j}$$

$$P_{max}^{cpu_i} = \frac{P_{max}^{host j} \times M_{curr}^{cpu_i}}{H_j}$$

Where:

- $P_{min}^{cpu_i}$: indicates the power deliver by a CPU of a host at idle state;
- $P_{max}^{cpu_i}$: indicates the power deliver by a CPU of a host at full state;
- $P_{curr}^{H_j}$: indicates the current capacity of the host in million instructions per second (MIPS);
- $P_{max}^{H_j}$: indicates the maximum capability of the host in MIPS;
- $P_{curr}^{cpu_i}$: indicates the current capacity of the CPU in MIPS;

- $P_{max}^{VM_k}$: indicates the maximum capacity of the VM in MIPS.

This measurement permits to find the best green host at a given time t, by considering every current CPU frequency.

In [51], researchers have explained a method of estimating energy consumption. So, the power model is:

$$E = \min(E_i) + CPUUsage \times (\max(E_i) - \min(E_i)) \quad (19)$$

Where:

- E_i : represents the amount of energy expended by a host;
- $CPUUsage$: indicates the average of mutual CPU usage among VMs running in the host;
- $\max(E_i)$: represents the maximum amount of energy consumed by the host in his maximum load;
- $\min(E_i)$: indicates the minimum energy expended by the host in his idle state [52].

Some researchers [53] have used the same previous formula, except that the symbols adopted are different. So, the calculation of the power consumption of the data center at time (t) is as follows:

$$P_{dc}(t) = \sum_{i=1}^{N_s} P_i(t) \quad (20)$$

For the calculation of power utilization of servers, according to [54], the power use of the CPU is taken in consideration, since this is the unit that exposes the higher change in power utilization concerning its use ratio.

Thus, for each server i at a time (t), the CPU use percentage $U_i(t)$ is equal to:

$$U_i(t) = \sum_{j=1}^{N_{vm}} \sum_{k=1}^{N_c} U_{c(k,j,i)}(t) \quad (21)$$

Where:

$$U_{c(k,j,i)}(t)$$

represents the CPU use of a container k on a VM j a server i at a time (t).

Also, by considering that the power consumption of a server is almost equal to null when it is turned off, then, its consumed power P_u at any specific processor use U, can be modeled as follows:

$$P_i(t) = P_i^{idle} + (P_i^{max} - P_i^{idle}) \times U_i(t) \quad \text{when } N_{vm} > 0 \quad (22)$$

Also, as estimate the total energy use of data center, other researchers like in [55], have adopted the server power model already proposed by known researchers like in [56]. The server power consumption calculation model is presented as:

$$P_i^{server} = P_i^{idle} + \sum_{j=1}^{W_i} w_j u(VM_{i,j}) \times P_i^{dynamic} \quad \text{when } w_i > 0 \quad (23)$$

Where:

- P_i^{server} : refers to the composition of idle power and dynamic power. Knowing that the idle power is considered as fixed and the dynamic power is linear to the total CPU utilization of all the VMs on the server;
 - P_i^{idle} : refers to the power when the server i is in idle state;
 - $P_i^{dynamic}$: refers to the power when the server i is fully loaded;
 - $VM_{i,j}$: indicates the j VM on the i server;
 - $u(VM_{i,j})$: indicates the utilization of the j VM on the i server;
 - w_i : indicates the number of VMs referred to a server i .
- The utilization $u(VM_{i,j})$ is determined as the calculation of the sum of all the application utilization on the j VM as follows:

$$u(VM_{i,j}) = \sum_{c=1}^{A_j} u(App_c) \quad (24)$$

Where:

- c : indicates the component;
 - A_j : indicates the number of application component.
- Also, in [57], researchers have designed another power model which is expressed as:

$$P_u = (P_{max} - P_{idle})(2u - u^r) + P_{idle} \quad (25)$$

Where: r represents an adjustment constant to reduce the square error that should be attained on an experimental basis. In [58], the researchers relied on the explanations of other researchers. S. Srikantaiah, et al. [59] explained that the power utilization of the hosts is controlled by CPU, RAM, network interfaces, and disk storage. With an end goal to expand on the power model in formula (22), in [60], it is made to some degree refined power consumption model as follows:

$$P_x(t) = (P_{x-idle}) + (P_{x-full} - P_{x-idle}) + \alpha_x U_x(t)^{\beta_x} \quad (26)$$

Where:

- P_{x-idle} : denotes the power utilization of a server x at an idle loaded state;
- P_{x-full} : corresponds to the power utilization of a server x at a fully loaded state;
- α , and β : represents dependent constants of a server;
- U_x : refers to the CPU use of a server x at a time t .

Prominent distinction from the power model expressed in formula (22) is the sum of a temporal parameter and the feature of accounting numerous diverse servers.

Other researchers [57] clarified that CPU utilization is the essential origin of the host's energy consumption.

As per the examination [61], the power model linked to CPU usage is characterized in the following formula:

$$P(h) = K \times h^{max} + (1 - K) \times h^{max} \times h^u \quad (27)$$

and:

- h^{max} : indicates the max power of the host h when its CPU utilization is equal to 100%;
- K : represents the rate of power consumption of an idle host.

Also, since the CPU can change over the time, we can consider [58], that the CPU utilization of a host is a function time and thus, the full energy consumption of a host can be represented as:

$$EC = \int_{t_0}^{t_1} P(h^u(t)) dt \quad (28)$$

According to [62] the energy consumption is determined by an accurate energy model formalized as follows:

$$E_{total} = E_{servers} + E_{network} \quad (29)$$

Also:

$$E_{total} = \sum_{h=host}^n predict(h) + E_{network} \quad (30)$$

Where:

- E_{total} : indicates the sum of cloud energy consumption;
- $E_{servers}$: represents the energy consumption of physical servers/nodes;
- $E_{network}$: refers to the energy consumption of network equipments (routers, switches, etc.);
- $predict(h)$: designates the computed prediction function adopted to determine a server consumption according to its metrics (CPU, RAM, disk, etc.) that is computed by a defined programmed algorithm based on machine learning.

In [63], for estimating the actual energy consumption EnC in the cloud data center, researchers adopted the following formula (31):

$$EnC = EnC_{Datacenter} + EnC_{Transceivers} + EnC_{Memory} + EnC_{Extra} \quad (31)$$

Where

- $EnC_{Datacenter}$: indicates the energy consumption of the data center;
- $EnC_{Transceivers}$: indicates the energy consumption of all the switching equipment;
- EnC_{Memory} : refers to the energy consumption of the storage device;
- EnC_{Extra} : designates the energy consumption of other components, like the fans, the current conversion loss and more...

Also, another further disintegrated formula is adopted [63] to express the data center energy consumption of a data centers, t transceivers material and a centralized memory equipment:

$$EnC = d(EnC_{Processor} + EnC_{PrimaryStorage} + EnC_{SecondaryStorage} + EnC_{Motherboards} + EnC_{NetworkCards}) + t'(EnC_{Hardware} + EnC_{LANcards} + \sum_{f=0}^F d_{connectors,f} + Enc_f) + (EnC_{NetworkAnalysisServer} + EnC_{MemoryManager} + EnC_{NetworkAttachedStorageArrays}) + EnC_{Extra} \quad (32)$$

Where:

- $EnC_{Hardware}$: indicates the energy consumption by the transceiver;
- $EnC_{LANcards}$: related to the energy consumption of any active network LAN card;
- EnC_f : represents the energy consumption of a connector (port) running at a frequency f ;

Also, in the formula (32), just the last component appears to be dependent on the connection recurrence while other components, as $EnC_{Hardware}$ and $EnC_{LANcards}$ stay fixed for all the duration of transceiver activity. Consequently, $EnC_{Hardware}$ and $EnC_{LANcards}$ can be kept away by turning the transceiver off or making it into sleep mode.

Moreover, in [63], the energy consumption at a given time t $EnC_{i,t}(r)$ is defined as follows:

$$EnC_{i,t}(r) = q \times EnC_{max} + (1 - q) \times EnC_{max} \times ru \quad (33)$$

Where:

- EnC_{max} : indicates the maximum energy consumption when resource is entirely used;
- q : represents the division of energy consumed by idle resource;
- ru : related to the resource utilization.

6. Proposition of a simulated and completed formula for calculating energy consumption in cloud data centers

As already specified, the aim of the cloud supplier is to minimize the actual energy consumption. The cloud workload will run just when the current energy consumption signified as (E_{act}) is lower than the threshold value of energy consumption.

For a specific cloud workload, to measure the energy consumption of resources for execution of heterogeneous cloud workloads, the information on its energy consumption and processor use is adopted.

The energy model is conceived on the assumption that resource use has an immediate association with energy consumption [64] [65]. In fact, The energy consumption of the data center at time t is determined as follows (34):

$$E_{act}(t) = \sum_{i=1}^n E_u \quad (34)$$

Where:

$\sum_{i=1}^n E_u$: refers to the energy consumption estimation of the whole server n of a user u in a time interval (t).

Formulating the problem

So as to reduce the power consumption of the data center with N VMs and M servers, we represent the issue as:

$$\min(E_{act}(t) = \sum_{i=1}^n E_u) \quad (35)$$

As well as:

$$\min(\sum_{i=1}^n P_u) \quad (36)$$

Where:

$\sum_{i=1}^n P_u$: represents the total power consumption estimation of the whole server n of a user u in a time interval (t).

And by considering some constraints, including:

$$\sum_{j=1}^N U_{vm_{j,i}} \langle S_{(i,r)}, \forall i \in [1, N_s], \quad (37)$$

$$\forall r \in \{CPU, memory, disk, bandwidth\}$$

Where:

- $U_{vm_{j,i}}(t)$: represents the resource use of VM j on the whole server i at a time interval (t);
- $S_{(i,r)}$: represents the server i capability of a resource r .

Also, the measure of energy consumption ($E_{act}(t)$) of using resources can be defined and as the coming equation (38):

$$E_{act}(t) = E_{dc} + E_{cpu} + E_{mem} + E_{disk} + E_{net} + E_{ext} \quad (38)$$

Where:

- $E_{act}(t)$: the total energy consumption in a time interval (t);
- E_{dc} : represents the data center's energy consumption;
- E_{cpu} : represents the energy consumption of the CPU device;
- E_{mem} : expressed the energy consumption of the storage device;
- E_{disk} : indicates the energy consumption of the disk device;
- E_{net} : refers to the energy consumption of all the switching and network material (network cards, network connectors, etc);
- E_{ext} : is the energy consumption of other parts, including the iops, the current conversion loss, the connectors, ports, motherboards, transceivers, and so forth.

The above formula can be further disintegrated; by defining the module E_{ext} to have a complete measure of energy consumed in the cloud data centers.

Also, the resource use vary in time and its function of time is expressed as $RU_{i,t}$ in equation (39):

$$RU_{i,t} = \sum_{i=1}^n E_{act_{i,t}} \times ru_t \quad (39)$$

We consider that:

- $E_{act_{t,i}}$: indicates the sum of the energy consumption at a given time t , where n is the number of cloud workloads running at time t ;
- ru_t : represents the resource utilization at a given time (t). According to the studied researches, the simplest and most complete formula for calculating the actual energy consumption $E_{act}(t)$ of a resource ru_t at a given time t is expressed as (40):

$$E_{act}(t) = RU_{i,t} + (E_{max} - E_{min}) + E_{min} \quad (40)$$

Where:

- $E_{max}(t)$: represents the maximum energy consumption of server i at the peak load (fully utilized server e.g 99% utilization);
- E_{min} : represents the minimum energy consumption in the

active/idle mode (low utilized server e.g 1% utilization).

6. Synthesis and discussion

Data centers represent the foundation of the present cloud computing systems. Because of the expanding interest in energy by data centers, it is important to represent the huge measure of expended energy. Power modeling and prediction of data centers assume a crucial aspect for this aim.

Thus, it was important to make a precise, in-depth and large research related to current studies in energy consumption modeling in the data center environment.

This study has shown us that according to the decomposition of the layers at the data center level: hardware layer and software layer, we can see that this has produced a hierarchy at the power model level.

Also, we noticed that there is an extensive quantity of researches directed on the energy utilization modeling in lower levels of the data center scale, considerably fewer studies are effected in the elevated layers. This can be considered as a crucial restriction of the actual state-of-the-art of researches related to power modeling.

Moreover, the exactness, generality, and practicality of most of the power consumption models stay open. In view of the direction deduced by our investigation, we predict noteworthy raise in energy modeling and prediction studies for higher levels of data center systems very soon.

7. Conclusion and perspectives

Cloud computing is a figuring recent paradigm used for demonstrating powerfully virtualized resource as a service via the network. Present data centers have widely included several technologies including virtualization, migration, and consolidation to acknowledge cloud computing.

However, these cited advances have their performance disadvantages. Indeed, a trade-off between the performance and energy efficiency still exists.

And as the fast development of server amount and scale in the data center, the energy used by this data center which is specifically identified with the quantity of hosted servers and their workloads is turning into an incredible challenge. Thus, it is crucial to design and develop energy-efficient advancements for the data center.

In this work, we made an exhaustive scientific categorization of power modeling and prediction strategies by presenting different formulas for calculating energy consumption in cloud data centers. In fact, various diverse insights gained through this review were described. And the major aim is to deduct a simplified and completed formula for energy consumption estimation.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License



References

- J. Whitney and P. Delforge, "Data center efficiency assessment," Issue paper on NRDC (The Natural Resource Defense Council), 2014.
- C. G. C. Index, "Forecast and methodology, 2015-2020 white paper," Retrieved 1st June, 2016.
- J. M. Kaplan, W. Forrest, and N. Kindler, "Revolutionizing data center energy efficiency," Technical report, McKinsey & Company, Tech. Rep., 2008.
- Y. Hua and D. Feng, "Needle in a haystack: Cost-effective data analytics for real-time cloud sharing," in 2014 IEEE 22nd International Symposium of Quality of Service (IWQoS). IEEE, 2014, pp. 159–167.
- S. K. Tesfatsion, "Energy-efficient cloud computing: autonomic resource provisioning for datacenters," Ph.D. dissertation, Umeå University, 2018.
- C. Preist and P. Shabajee, "Energy use in the media cloud: Behaviour change, or technofix?" in Cloud Computing Technology and Science (CloudCom), 2010 IEEE Second International Conference on. IEEE, 2010, pp. 581–586.
- L. A. Barroso, J. Clidaras, and U. Holzle, "The datacenter as a computer: An introduction to the design of warehouse-scale machines," Synthesis lectures on computer architecture, vol. 8, no. 3, pp. 1–154, 2013.
- R. Buyya, C. Vecchiola, and S. T. Selvi, "Mastering cloud computing: foundations and applications programming," Newnes, 2013.
- R. Buyya, A. Beloglazov, and J. Abawajy, "Energy-efficient management of data center resources for cloud computing: a vision, architectural elements, and open challenges," arXiv preprint arXiv:1006.0308, 2010.
- L. Krug, M. Shackleton, and F. Saffre, "Understanding the environmental costs of fixed line networking," in Proceedings of the 5th international conference on Future energy systems. ACM, 2014, pp. 87–95.
- M. Dayarathna, Y. Wen, and R. Fan, "Data center energy consumption modeling: A survey," IEEE Communications Surveys & Tutorials, vol. 18, no. 1, pp. 732–794, 2016.
- M. Poess and R. O. Nambiar, "Energy cost, the key challenge of today's data centers: a power consumption analysis of tpe-c results," Proceedings of the VLDB Endowment, vol. 1, no. 2, pp. 1229–1240, 2008.
- Y. Gao, H. Guan, Z. Qi, B. Wang, and L. Liu, "Quality of service aware power management for virtualized data centers," Journal of Systems Architecture, vol. 59, no. 4-5, pp. 245–259, 2013.
- S. Rivoire, M. A. Shah, P. Ranganathan, C. Kozyrakis, and J. Meza, "Models and metrics to enable energy-efficiency optimizations," Computer, vol. 40, no. 12, 2007.
- K. Bilal, S. U. R. Malik, S. U. Khan, and A. Y. Zomaya, "Trends and challenges in cloud datacenters," IEEE cloud computing, vol. 1, no. 1, pp. 10–20, 2014.
- B. Whitehead, D. Andrews, A. Shah, and G. Maidment, "Assessing the environmental impact of data centres part 1: Background, energy use and metrics," Building and Environment, vol. 82, pp. 151–159, 2014.
- V. Mathew, R. K. Sitaraman, and P. Shenoy, "Energy-aware load balancing in content delivery networks," in INFOCOM, 2012 Proceedings IEEE. IEEE, 2012, pp. 954–962.
- P. Corcoran and A. Andrae, "Emerging trends in electricity consumption for consumer ict," National University of Ireland, Galway, Connacht, Ireland, Tech. Rep, 2013.
- J. Kooimey, "Growth in data center electricity use 2005 to 2010," A report by Analytical Press, completed at the request of The New York Times, vol. 9, 2011.
- W. Van Heddeghem, S. Lambert, B. Lannoo, D. Colle, M. Pickavet, and P. Demeester, "Trends in worldwide ict electricity consumption from 2007 to 2012," Computer Communications, vol. 50, pp. 64–76, 2014.
- D. J. Brown and C. Reams, "Toward energy-efficient computing," Queue, vol. 8, no. 2, p. 30, 2010.
- N. M. Seel, "Encyclopedia of the Sciences of Learning," Springer Science & Business Media, 2011.
- S. M. Rivoire, "Models and metrics for energy-efficient computer systems," Stanford University, 2008.
- M. von dem Berge, G. Da Costa, A. Kopecki, A. Oleksiak, J.-M. Pierson, T. Piontek, E. Volk, and S. Wesner, "Modeling and simulation of data center energy-efficiency in coolemall," in International Workshop on Energy Efficient Data Centers. Springer, 2012, pp. 25–36.
- A. Floratou, F. Bertsch, J. M. Patel, and G. Laskaris, "Towards building wind tunnels for data center design," Proceedings of the VLDB Endowment, vol. 7, no. 9, pp. 781–784, 2014.
- D. C. Kilper, G. Atkinson, S. K. Korotky, S. Goyal, P. Vetter, D. Suvakovic, and O. Blume, "Power trends in communication networks," IEEE Journal of Selected Topics in Quantum Electronics, vol. 17, no. 2, pp. 275–284, 2011.
- H. Xu and B. Li, "Reducing electricity demand charge for data centers with partial execution," in Proceedings of the 5th international conference on Future energy systems. ACM, 2014, pp. 51–61.
- D. Wang, C. Ren, S. Govindan, A. Sivasubramaniam, B. Uргаonkar, A. Kansal, and K. Vaid, "Ace: Abstracting, characterizing and exploiting datacenter power demands," in 2013 IEEE International Symposium on

- Workload Characterization (IISWC). IEEE, 2013, pp. 44–55.
29. J. J. Evans, "On performance and energy management in high performance computing systems," in *Parallel Processing Workshops (ICPPW)*, 2010 39th International Conference on. IEEE, 2010, pp. 445–452.
 30. S. Roy, A. Rudra, and A. Verma, "An energy complexity model for algorithms," in *Proceedings of the 4th conference on Innovations in Theoretical Computer Science*. ACM, 2013, pp. 283–304.
 31. B. M. Tudor and Y. M. Teo, "On understanding the energy consumption of arm-based multicore servers," in *ACM SIGMETRICS Performance Evaluation Review*, vol. 41, no. 1. ACM, 2013, pp. 267–278.
 32. A. Kansal, F. Zhao, J. Liu, N. Kothari, and A. A. Bhattacharya, "Virtual machine power metering and provisioning," in *Proceedings of the 1st ACM symposium on Cloud computing*. ACM, 2010, pp. 39–50.
 33. R. Ge, X. Feng, and K. W. Cameron, "Modeling and evaluating energy-performance efficiency of parallel processing on multicore based power aware systems," in *Parallel & Distributed Processing, 2009. IPDPS 2009. IEEE International Symposium on*. IEEE, 2009, pp. 1–8.
 34. S. L. Song, K. Barker, and D. Kerbyson, "Unified performance and power modeling of scientific workloads," in *Proceedings of the 1st International Workshop on Energy Efficient Supercomputing*. ACM, 2013, p. 4.
 35. A. Lewis, J. Simon, and N.-F. Tzeng, "Chaotic attractor prediction for server run-time energy consumption," in *Proceedings of the 2010 international conference on Power aware computing and systems*. USENIX Association, 2010, pp. 1–16.
 36. A. W. Lewis, N.-F. Tzeng, and S. Ghosh, "Runtime energy consumption estimation for server workloads based on chaotic time-series approximation," *ACM Transactions on Architecture and Code Optimization (TACO)*, vol. 9, no. 3, p. 15, 2012.
 37. V. Perumal and S. Subbiah, "Power-conservative server consolidation based resource management in cloud," *International Journal of Network Management*, vol. 24, no. 6, pp. 415–432, 2014.
 38. A. Chatzipapas, D. Peditatitakis, C. Rotsos, V. Mancuso, J. Crowcroft, and A. Moore, "Challenge: Resolving data center power bill disputes: The energy-performance trade-offs of consolidation," in *Proceedings of the 2015 ACM Sixth International Conference on Future Energy Systems*. ACM, 2015, pp. 89–94.
 39. I. Alan, E. Arslan, and T. Kosar, "Energy-aware data transfer tuning," in *Cluster, Cloud and Grid Computing (CCGrid)*, 2014 14th IEEE/ACM International Symposium on. IEEE, 2014, pp. 626–634.
 40. A. Beloglazov, R. Buyya, Y. C. Lee, and A. Zomaya, "A taxonomy and survey of energy-efficient data centers and cloud computing systems," in *Advances in computers*. Elsevier, 2011, vol. 82, pp. 47–111.
 41. F. Chen, J. Grundy, Y. Yang, J.-G. Schneider, and Q. He, "Experimental analysis of task-based energy consumption in cloud computing systems," in *Proceedings of the 4th ACM/SPEC International Conference on Performance Engineering*. ACM, 2013, pp. 295–306.
 42. P. Xiao, Z. Hu, D. Liu, G. Yan, and X. Qu, "Virtual machine power measuring technique with bounded error in cloud environments," *Journal of Network and Computer Applications*, vol. 36, no. 2, pp. 818–828, 2013.
 43. G. Dhiman, K. Mihic, and T. Rosing, "A system for online power prediction in virtualized environments using gaussian mixture models," in *Proceedings of the 47th Design Automation Conference*. ACM, 2010, pp. 807–812.
 44. R. Lent, "A model for network server performance and power consumption," *Sustainable Computing: Informatics and Systems*, vol. 3, no. 2, pp. 80–93, 2013.
 45. A.-C. Orgerie, L. Lef'evre, and I. Gu'erin-Lassous, "Energy-efficient bandwidth reservation for bulk data transfers in dedicated wired networks," *The Journal of Supercomputing*, vol. 62, no. 3, pp. 1139–1166, 2012.
 46. Y. Li, Y. Wang, B. Yin, and L. Guan, "An online power metering model for cloud environment," in *Network Computing and Applications (NCA)*, 2012 11th IEEE International Symposium on. IEEE, 2012, pp. 175–180.
 47. X. Xu, K. Teramoto, A. Morales, and H. H. Huang, "Dual: Reliabilityaware power management in data centers," in *2013 13th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing*. IEEE, 2013, pp. 530–537.
 48. Z. Zhang, Q. Guan, and S. Fu, "An adaptive power management framework for autonomic resource configuration in cloud computing infrastructures," in *Performance Computing and Communications Conference (IPCCC)*, 2012 IEEE 31st International. IEEE, 2012, pp. 51–60.
 49. B. Bermejo, C. Guerrero, I. Lera, and C. Juiz, "Cloud resource management to improve energy efficiency based on local nodes optimizations," *Procedia Computer Science*, vol. 83, pp. 878–885, 2016.
 50. T. Guerout and M. B. Alaya, "Autonomic energy-aware tasks scheduling," in *Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, 2013 IEEE 22nd International Workshop on. IEEE, 2013, pp. 119–124.
 51. M. M. Alansari and B. Bordbar, "Modelling and analysis of migration policies for autonomic management of energy consumption in cloud via petri-nets," in *2014 International Conference on Cloud and Autonomic Computing (ICCAC)*. IEEE, 2014, pp. 121–130.
 52. D. Borgetto, M. Maurer, G. Da-Costa, J.-M. Pierson, and I. Brandic, "Energy-efficient and sla-aware management of iaas clouds," in *Proceedings of the 3rd International Conference on Future Energy Systems: Where Energy, Computing and Communication Meet*. ACM, 2012, p. 25.
 53. S. F. Piraghaj, A. V. Dastjerdi, R. N. Calheiros, and R. Buyya, "A framework and algorithm for energy efficient container consolidation in cloud data centers," in *Data Science and Data Intensive Systems (DSDIS)*, 2015 IEEE International Conference on. IEEE, 2015, pp. 368–375.
 54. M. Blackburn and G. Grid, "Five ways to reduce data center server power consumption," *The Green Grid*, vol. 42, p. 12, 2008.
 55. M. Xu, A. V. Dastjerdi, and R. Buyya, "Energy efficient scheduling of cloud application components with brownout," *IEEE Transactions on Sustainable Computing*, vol. 1, no. 2, pp. 40–53, 2016.
 56. K. Zheng, X. Wang, L. Li, and X. Wang, "Joint power optimization of data center network and servers with correlation analysis," in *INFOCOM, 2014 Proceedings IEEE*. IEEE, 2014, pp. 2598–2606.
 57. X. Fan, W.-D. Weber, and L. A. Barroso, "Power provisioning for a warehouse-sized computer," in *ACM SIGARCH computer architecture news*, vol. 35, no. 2. ACM, 2007, pp. 13–23.
 58. Y. Liu, X. Sun, W. Wei, and W. Jing, "Enhancing energy-efficient and qos dynamic virtual machine consolidation method in cloud environment," *IEEE Access*, 2018.
 59. S. Srikantaiah, A. Kansal, and F. Zhao, "Energy aware consolidation for cloud computing," 2008.
 60. C.-J. Tang and M.-R. Dai, "Dynamic computing resource adjustment for enhancing energy efficiency of cloud service data centers," in *System Integration (SI)*, 2011 IEEE/SICE International Symposium on. IEEE, 2011, pp. 1159–1164.
 61. S. K. Garg, A. N. Toosi, S. K. Gopalaiyengar, and R. Buyya, "Slabased virtual machine management for heterogeneous workloads in a cloud datacenter," *Journal of Network and Computer Applications*, vol. 45, pp. 108–120, 2014.
 62. E. Outin, J.-E. Dartois, O. Barais, and J.-L. Pizat, "Enhancing cloud energy models for optimizing datacenters efficiency," in *Cloud and Autonomic Computing (ICCAC)*, 2015 International Conference on. IEEE, 2015, pp. 93–100.
 63. S. Singh, I. Chana, M. Singh, and R. Buyya, "Soccer: self-optimization of energy-efficient cloud resources," *Cluster Computing*, vol. 19, no. 4, pp. 1787–1800, 2016.
 64. S. Singh and I. Chana, "Earth: Energy-aware autonomic resource scheduling in cloud computing," *Journal of Intelligent & Fuzzy Systems*, vol. 30, no. 3, pp. 1581–1600, 2016.
 65. S. Sukhpal and C. Inderverer, "Resource provisioning and scheduling in clouds: Qos perspective," *The Journal of Supercomputing*, vol. 72, no. 3, pp. 926–960, 2016.
 66. W. L. Bircher and L. K. John, "Complete system power estimation using processor performance events," *IEEE Transactions on Computers*, vol. 61, no. 4, pp. 563–577, 2011.
 67. D. Economou and S. Rivoire and C. Kozyrakis and P. Ranganathan, "Full-system power analysis and modeling for server environments," *International Symposium on Computer Architecture-IEEE*, 2006.
 68. R. Rodrigues A. Annamalai I. Koren and S. Kundu, "A study on the use of performance counters to estimate power in microprocessors," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 60, no. 12, pp. 882–886, 2013.
 69. K. Singh M. Bhaduria and S. A. McKee, "Real time power estimation and thread scheduling via performance counters," *ACM SIGARCH Computer Architecture News*, vol. 37, no. 2, pp. 46–55, 2009.
 70. R. Joseph and M. Martonosi, "Run-time power estimation in high performance microprocessors," in *ISLPED'01: Proceedings of the 2001 International Symposium on Low Power Electronics and Design (IEEE Cat. No. 01TH8581)* (pp. 135–140), 2001.
 71. C. Gilberto and M. Margaret, "Power prediction for intel XScale processors using performance monitoring unit events power prediction for intel XScale processors using performance monitoring unit events," in *ISLPED*, vol. 5, pp. 8–10, 2005.
 72. R. Berrendorf and H. Ziegler, "PCL—the Performance Counter Library: A Common Interface to Access Hardware Performance Counters on Microprocessors," Version 1.3, 1998.