Trends and Effects of Energy Proportionality on Server Provisioning in Data Centers

Georgios Varsamopoulos, Zahra Abbasi and Sandeep K. S. Gupta
Impact Laboratory (http://impact.asu.edu/)
Arizona State University
{georgios.varsamopoulos,zahra.abbasi,sandeep.gupta}@asu.edu

***Abstract*—Cloud is the state-of-the-art back-end infrastructure for most large-scale web services. This paper studies what effect energy proportionality has on the energy savings of cloud data center management, under various equipment compositions and power densities. Our findings show that although it is a common expectation that improved energy proportionality should diminish the benefits of power management's server provisioning, this is not true in all cases. Results show that equipping server provisioning with thermal awareness can keep it as a useful technique when the data center exhibits power consumption heterogeneity and non-uniform heat recirculation phenomena.**

***Index Terms*—Energy proportionality; data centers; server provisioning; thermal awareness.**

I. Introduction

A cloud computing or a cluster computing application relies on a back-end infrastructure of hundreds or thousands of servers, located in one or more *data centers*. Such Internet data centers have seen a tremendous population and energy consumption growth in the past decade, with the energy efficiency being a major research topic.

Up until now, energy efficiency issues have been largely addressed by power management solutions, which suspend servers that are not utilized. This is also known as *active server set provisioning*, and it has been shown to successfully to address the problem of idle server power consumption [1]–[4]. It however does not address the energy wastage when servers are each partially utilized.

Recently, the principle of energy-proportional computing has been proposed [5], which suggests that systems should consume power in proportion to their utilization level. The idea is to address the observation that in most applications, servers are used at 10-50% of their peak computing capacity, but their power consumption at those levels is comparable to their peak [5].

Although energy proportionality was proposed as a measure orthogonal to active server set provisioning, it is expected to reduce the energy-saving benefits of server provisioning because it significantly reduces the idle power consumption. Energy proportionality, as proposed in the literature, is to address the energy wastage at the partially utilized severs; energy-proportional computing systems would spend only as much energy as the given load. However, if the idle servers were not turned off, they would consume no power under ideal energy proportionality.

The main objective of this paper is to investigate whether increased (and eventually ideal) energy proportionality will render server provisioning an unnecessary technique. This is done by performing a simulation-based study of the energy consumption of various thermal-aware and non-thermal-aware server provisioning schemes under various energy proportionality cases. Results in this paper suggest that *server provisioning will still offer energy savings over no provisioning, although significantly reduced, under energy-proportional computing in data centers that exhibit equipment power heterogeneity and heat recirculation.*

A. Overview of results and contributions

This paper proposes two quantitative metrics for measuring energy proportionality, namely idle-to-power ratio (IPR) and linear deviation ratio (LDR). The first metric is a measure of the power range, whereas the second is a measure of linearity. Applying these metrics to power curves published by the SPECpower_ssj2008 webpage[1], two historical trends are yielded (2007–2010): IPR shows an improvement, meaning that the power range is becoming larger, more "proportional", whereas the LDR is worsening, meaning that power curves deviate from being linear.

The second result of the paper is a study of how the energy savings of server provisioning and is affected by an improving IPR of the server systems. The results of the study show that:

- Although the benefits of server provisioning reduce with an increased energy proportionality, they are above zero in heterogeneous data centers. This is possible if server provisioning is done with thermal-awareness and the workload distribution is done with load balancing.
- Under full energy proportionality, almost all of the energy-savings will come from thermal-aware workload distribution. This means that *to completely phase-out server provisioning, energy proportionality alone is not enough; energy (thermal) awareness must also to be employed.*

The study in this paper also assesses the energy savings of various workload distribution methods and shows that thermal-

This work has been partly funded by NSF CRI grant #0855527 and by CNS grant #0834797. A short version of the paper focusing on the metrics section appeared in the "GreenCom 2010" ICPP Workshop, Sep 13, 2010.

[1] http://www.spec.org/power_ssj2008/results/

aware workload distribution that optimizes for the total energy consumption of the data center (which is modeled as the sum of computing and cooling energy consumption) has generally increasing energy savings over a workload balancing approach.

B. Paper organization

The investigation approach in this paper is divided in four steps: first, it reviews the current state of data centers and their dominant physical aspects (§II), those being power density and heterogeneity, and the software architecture, which consists two tiers, one of resource management and one of workload management. Second, it introduces the IPR and LDR metrics, which are used to classify how "energy-proportional" a computing system is (§ III); it then it investigates the technological trends of energy proportionality (§ III-A). Third, it compares the energy savings of server provisioning (with respect to no sever provisioning) and the savings of thermal-aware workload distribution over equal load balancing in a cloud computing data center with respect to IPR and LDR (§IV).

II. Data centers: current state

A. Physical aspects

A typical data center has its equipment placed on a raised floor in the so-called "hot aisle / cold aisle" layout. The raised floor in the cold aisles features perforations which allow cool air to enter the room; perforations or other contraptions above the hot aisles gather the hot air, which is passed to the computer room air conditioner (CRAC).

The computing equipment itself is in rack-mount (for older technologies) or blade server (for newer technologies) organization. Server cabinets (a.k.a. racks) contain up to five server enclosures (a.k.a. chassis); part of the cabinet space is filled with power distribution, networking and storage equipment. Each chassis contains eight to sixteen blade servers. Blade-based organization allows for ***high power density***, which can reach up to 5000 W/ft^2 by year 2014 [6], which means 30-35 KW per rack. Power density is a significant parameter in data center design, because it largely affects the cooling design (and the overall data center power consumption).

Another aspect of data centers is their ***equipment heterogeneity***. Most data centers are partially upgraded on a 2-year or 3-year cycle. For 5-year old data centers (most of data centers are at least that old), this means several generations of equipment, thus resulting in heterogeneity, where systems have different computing capacity, power rating, and corresponding computing energy efficiency.

These aspects, i.e. the power density and heterogeneity, will be used as parameters in the simulation-based study of §IV.

B. Related work on addressing the energy waste.

The basic premise for saving energy is the dynamic variability of traffic, i.e. its intensity [7]–[9]. This dynamic variation originates from the variability of file sizes and the collective user behavior. Saving energy is done by dynamically adjusting the computing capacity to match the traffic intensity.

Historically, there have been three research directions toward saving energy in data centers: i) suspending unnecessary systems, ii) dynamic frequency and voltage scaling, and iii) energy-aware management of workload.

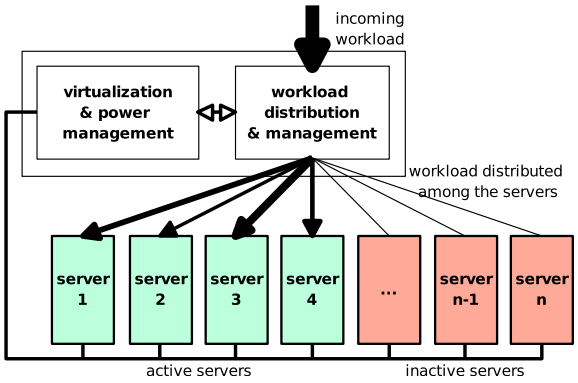


Fig. 1. Generic management architecture for cloud computing. It is divided into two tiers: the resource (power, virtualization) management and the workload management.

1) Suspending or turning off systems: Suspending unnecessary systems has great potential for energy savings especially when the incoming workload is much less than the capacity of the entire data center. This idea evolved into the concept of power management through active server set provisioning, which involves estimating how many servers are needed to service the incoming workload [1], [3], [4]; the excess servers are suspended. The determination of the active server set relies on good prediction of the upcoming workload and the data center's ability to service it, and there have been several efforts on those issues [1]–[4].

2) DVFS: Dynamic frequency and voltage scaling is a technique of dynamically changing the operating frequency and voltage of system components with respect to the computational demands. Although DVFS is still an active area of research [9]–[11], firmware-level technologies of DVFS are introduced to production systems, also known as *CPU throttling*. Throttling has become a common feature in modern CPUs (e.g. AMD's *Cool'n'Quiet* and *PowerNow!* and Intel's *SpeedStep* technologies); the ACPI standard, as of version 2.0, specifies the so-called performance states (P-states). CPU throttling forms the dominant technological base for implementing energy-proportional computing at single-system level.

3) Energy-aware approaches: Model-based energy-aware scheduling and workload management approaches have been proposed to allocate the workload in such a way so as to reduce the combined computing and cooling energy costs. Power-aware approaches try to reduce the computing energy consumption by selecting energy-efficient servers for a given workload, whereas thermal-aware approaches take into account the thermal-impact of workload for workload distribution [12]–[17], [17].

Part of this paper's objectives is to assess whether increased energy proportionality would reduce the importance of thermal-aware approaches.

C. Software management architecture

The management software is responsible for various functions, including power management, resource provisioning,

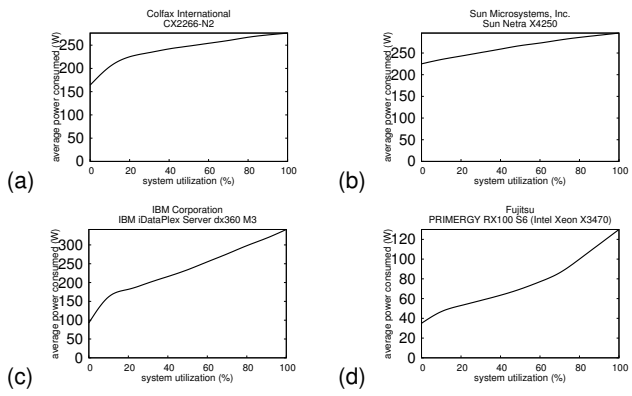


Fig. 2. Example power curves of various systems as published at SPECpower_ssj2008[1]

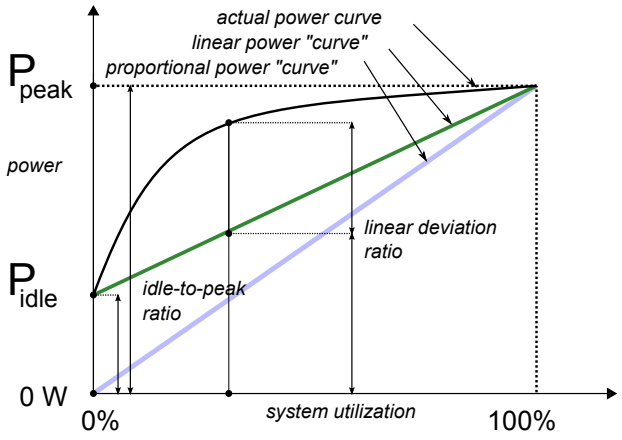


Fig. 3. Two proposed metrics of energy proportionality: idle-to-peak ratio (IPR) and linear deviation ratio of power (LDR).

workload management and status monitoring. The software management architecture for clouds can be organized into two generic tiers: one tier that is responsible for ***resource, virtualization and power management***, and another tier which is responsible for ***management and distribution of workload*** among the servers. Figure 1 gives a graphical representation of the architecture. This organization, although it may look simplistic, abstracts the predominant functionality of the management software, which is to manage systems (Tier 1) and to manage workload (Tier 2).

One of the resource management responsibilities is to estimate how many servers are needed to service the incoming workload. In the case of multiple serviced applications, this layer decides on how the servers are to be split among the applications. Virtualization technologies offer great flexibility to this purpose.

Workload management and distribution involves splitting incoming workload among the allocated servers, always with respect to SLA requirements.

III. Energy proportionality metrics

Although energy proportionality is a fundamental engineering objective, as envisioned in [5], [18] , there is lack of proper quantitative metrics to classify "how energy proportional" a system is. This section proposes two metrics that quantify the energy proportionality of a system.

Contemporary computing systems are far from being energy proportional; their idle power consumption is nowhere close to zero and their power curve is not a straight line. Figure 2 demonstrates these properties; it was created using data from the SPECpower_ssj2008 published results web page[1]. The figure shows that the power consumption curves start at a point above zero, and do not follow a straight line to maximum. Notice also that the biggest "deviation" from the straight line between the idle power (P_{idle}) and the peak power (P_{peak}) happens before 50%.

From the above it can be concluded that, in order to measure the energy proportionality of a system, one needs to measure how close to the origin the power curve starts and how close to linear the it is. In other words, two metrics are needed: one that measures the "*range*" and one the "*linearity*" (the second criterion is just as important; §III-A below shows that computing systems with greater range tend to be less linear).

For the power "range" aspect, we propose the *idle-to-peak power ratio* (IPR), which is defined as the ratio of the power consumption at 0% utilization over the power consumption at 100% utilization (see Figure 3),

$$\text{IPR} = P_{\text{idle}}/P_{\text{peak}}. \tag{1}$$

Lower IPR values denote a more energy-proportional system. IPR has an advantage over using the absolute idle power as a metric, because it is normalized over the *dynamic range* of power consumption of a system, thus favoring systems that have a larger distance between idle and peak power. Also, since it is normalized, it can be used for direct comparison among systems of different power consumption magnitude.

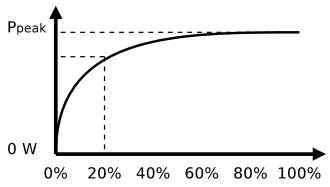
For the "linearity" aspect of the power consumption, we proposed the *linear deviation ratio* (LDR), which is defined as the maximum ratio of the actual power consumption's difference from the *hypothetical* linear power consumption P_{idle} at 0% to P_{peak} at 100% utilization, over the hypothetical linear consumption (see Figure 3):

$$\text{LDR} = \max_{u}^{|\cdot|} \frac{P(u) - \left(\left(P_{\text{peak}} - P_{\text{idle}}\right) u + P_{\text{idle}}\right)}{\left(p_{\text{peak}} - P_{\text{idle}}\right) u + P_{\text{idle}}}, \tag{2}$$

where $\max^{|\cdot|}$ is the maximum by absolute-value comparison which retains the sign of the maximum value. This is because we want LDR to retain the sign of the maximum deviation.

Lower LDR values denote a more linear system. Negative LDR values denote a power curve that is under the straight line. Positive LDR values denote a power curve that is over the straight line. LDR is also normalized, so it can be used for direct comparison. Note that LDR penalizes deviations that occur closer to the lower end of the curve. This is in accordance to the observation in Figure 2, showing that the bigger deviations happen below 50% utilization. Fig. 2(a) has high IPR and LDR, Fig. 2(b) has high IPR and low LDR, Fig. 2(c) has low IPR and high LDR, and Fig. 2(d) has low IPR and low LDR.

A low IPR value alone does not guarantee energy-proportional behavior. For example, consider the following hypothetical power curve:



In the above curve, IPR is zero while LDR is extremely high; the power consumption reaches 80% of the peak at 20% utilization. Therefore, for proper energy proportional behavior, a system must have ***both*** its IPR and LDR approach zero.

A. Energy proportionality trends in production systems

The study in this section applies the IPR and LDR metrics to power benchmarking results in order to yield historical trends of technology toward energy proportionality. The published results of SPECpower_ssj2008[1] were chosen as the data. The study includes all entries since 2007 (139 entries). IPR was calculated using the "average watts at idle" and "average watts at 100% utilization" entries of each system. LDR was calculated using all eleven average wattage recordings (0%, 10%,...,100%).

Figure 4 shows the historic trends of IPR and LDR for systems from early 2007 until July 2010. It was created using the IPR values over the "Hardware availability" entry of the database, which is used in this paper as the hardware release date. There is a clear technological trade-off between IPR and LDR, where modern systems tend to have a significantly improved IPR at the cost of increased LDR. The figure also shows two LDR trends, one positive LDR trend which is dominant, and a lesser negative LDR trend. These two trends make the scatter plot take a λ-shaped form. It is an important issue for the computing industry to address the increasingly large LDR values.

IV. Energy proportionality and energy savings of cloud management

This section studies how power density, heterogeneity and energy proportionality affect the energy efficiency of a data center. To that end, we build graphs of energy savings with IPR (Eq. 1) on the horizontal axis, for various power densities, for homogeneous and heterogeneous data centers. The range spans from IPR of 1 on the left ("energy-constant") to 0 on the right ("fully" energy-proportional). Specifically, we study: i) the energy savings of power management by server provisioning over no power management, and ii) the energy savings of thermal-aware workload distribution over equal load balancing.

A. Data center energy consumption model

This subsection describes the energy consumption model of a data center from a holistic, thermodynamic point of view. It is an overview of the model in [13], [14]. According to this model, the total power consumption of a data center is the sum of the cooling energy and computing energy:

$$P_{\text{total}} = P_{\text{comp}} + P_{\text{AC}}. \tag{3}$$

The energy consumed by the CRAC is dependent on its *coefficient of performance* (CoP), which is the ratio of the heat removed over the work required to remove that heat; a higher CoP means more efficient cooling, and usually the higher the required operating temperatures the better the CoP.

$$P_{\text{total}} = P_{\text{comp}} + \frac{P_{\text{comp}}}{\text{CoP}(T_{\text{AC}})} = \left(1 + \frac{1}{\text{CoP}(T_{\text{AC}})}\right) P_{\text{comp}}. \tag{4}$$

The highest CRAC output temperature (T_{AC}) is limited by the servers' air inlet *redline temperature*, i.e. the maximum temperature of cool air that has to enter the server's air inlet, offset by the *maximum temperature rise* caused by heat that is recirculated in the data center room into the servers' air inlets:

$$T_{\text{AC,max}} = T_{\text{red}} - \max\{T_{\text{rise}}\}. \tag{5}$$

The heat recirculation in a data center room is modeled as a matrix $\boldsymbol{D}=\{d_{ij}\}_{N\times N}$ of coefficients, where N is the number of distinct computing elements (depending on the desired granularity, these can be servers, enclosures, or racks). In this case element d_{ij} of this matrix is the coefficient of heat that is distributed from chassis i to chassis j [13] (this matrix also converts heat to temperature). Let $\boldsymbol{p}_{\text{comp}}$ be the current computing power consumption vector, where each vector element $p_{\text{comp},i}$ denotes the power consumption of each server, then the maximum temperature rise is given by:

$$\max\{T_{\text{rise}}\} = \max\{\boldsymbol{Dp}\}. \tag{6}$$

Then, the total power consumption is expressed as:

$$P_{\text{total}} = \left(1 + \frac{1}{\text{CoP}\left(T_{\text{red}} - \max\{\boldsymbol{Dp}\}\right)}\right) \sum p_{\text{comp},i}, \tag{7}$$

Computing power can be calculated through CPU utilization which is an indication of total power consumption of a typical server [19]. The total power consumption of servers having CPU utilization of u_i each can be written as $\boldsymbol{p}(\boldsymbol{u})$, where $\boldsymbol{u}$ represents the utilization vector of the servers. Note that, *since the utilization vector* $\boldsymbol{u}$ *directly depends on the workload distribution to the servers, the energy consumption also depends on the workload distribution.*

1) Energy-saving model for Internet data centers: Power management solutions practically zero the parameters a and w for the systems they suspend. That means that if an active server set scheme is used, then the summation in Eq. 7 is reduced to the active servers. If the active server set is denoted as $\mathcal{S}$, then Eq. 7 is translated to:

$$P_{\text{total}} = \left(1 + \frac{1}{\text{CoP}\left(T^{\text{red}} - \max_{\mathcal{S}}\{\boldsymbol{Dp}_{\mathcal{S}}(\boldsymbol{u}_{\mathcal{S}})\}\right)}\right) \sum_{i\in\mathcal{S}} p_i(u_i). \tag{8}$$

Workload management solutions distribute the workload and yield a utilization u_i on each server i for the level of workload λ_i (i.e. arrival rate) that it is assigned:

$$u_i = c_i\lambda_i,$$

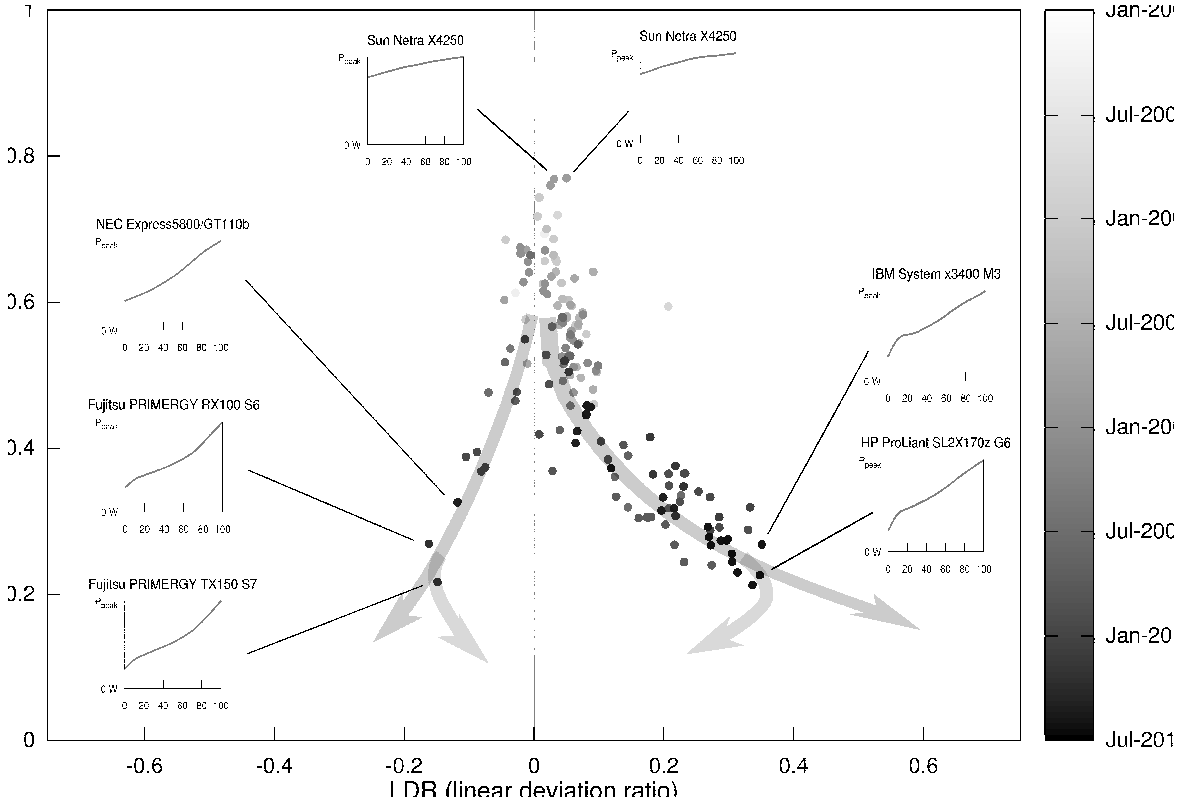


Fig. 4. Scatter plot of (LDR,IPR) pairs for each benchmark entry by SPECpower2008[1]. Each dot is grayscale-mapped to the hardware release date. Power curves for selected entries have been inserted in the figure. The plot distinctively shows the divergence trend of LDR, as highlighted by the thick arrows. It is an imperative challenge to reverse the trend of LDR and make it converge toward 0, along with IPR.

where c_i is the conversion coefficient that denotes the utilization level per workload unit, i.e. per arrival rate unit.

To compare two schemes, one only needs to substitute their values in Eq. 7 and compute the savings with the following formula:

$$\text{savings} = \frac{E_{\text{total, target_scheme}} - E_{\text{total, reference_scheme}}}{E_{\text{total, reference_scheme}}} \qquad (9)$$

B. The selected server provisioning and workload distribution approaches

This subsection describes the selected thermal-aware server provisioning and workload distribution approaches used in the following evaluation subsection.

1) Server provisioning approach: In previous work, thermal-aware server provisioning (TASP) has been shown to yield the most energy savings, and it is shown to be considerably better than simply power-aware schemes [20]. Therefore, TASP is chosen as the scheme to be studied in this paper.

The logic behind TASP is to do a periodic active server set selection. The selection period is referred to as *epoch*. The objective of TASP is to select that active server set that will be able to service the estimated traffic peak within the epoch (this is regarded as the *performance constraint*) which will help minimize Eq. 8. Pivotal role in the set size is the per-server *utilization threshold* $u_{\text{threshold}}$, which denotes the maximum utilization level of a server for which the SLA-posed performance is met; if a server is utilized above $u_{\text{threshold}}$ then it will most probably fail the SLA requirement. This utilization-based description of performance is evidenced in [4], [20].

The TASP problem is formulated as a non-linear binary minimax optimization problem, in which if the servers were to be used at their maximum performance-guarantying level $u_{\text{threshold}}$

> **TASP:** Given the incoming traffic intensity λ, select the active server set $\mathcal{S}$ such that if the active servers are utilized at their maximum performance threshold $u_{\text{threshold},i}$ each, Eq 8 is minimized.

This problem can be solved using various numeric methods, e.g. sequential quadratic programming (accompanied by a solution discretization step) or branch-and-bound. See [20] for details of solution approaches. For the scope of this paper, the specifics of the algorithms are not as important as the optimization objectives that characterize the methods.

2) Workload distribution approaches: The study considers the following workload distribution heuristics, taken from [20], which stochastically split the incoming requests among the active servers according on a probability vector $\boldsymbol{\pi} = \langle \pi_i \rangle$, where π_i is the probability of a request being assigned to active server i.

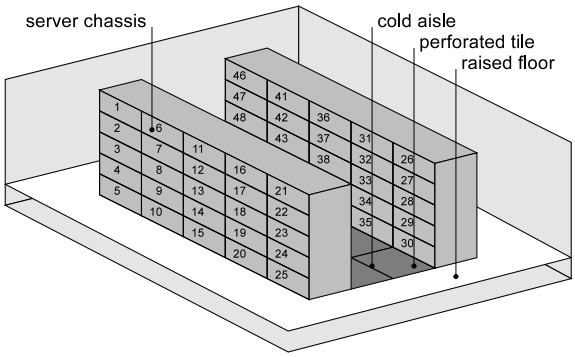


Fig. 5. The physical layout of the data center used in the study. It is a two-row blade-server data center room. The chassis are numbered in a top-bottom, horseshoe fashion.

Load balancing: this workload distribution approach tries to evenly balance[2] the load, i.e. $\lambda_i = \frac{c_i}{\sum_j c_j}\lambda$. In this approach, $\boldsymbol{\pi} = \left\langle \frac{c_i}{\sum_j c_j} \right\rangle$. If the servers are all computationally equal, then $\boldsymbol{\pi} = \langle 1/|\mathcal{S}| \rangle$.

Total energy minimization: this workload distribution approach divides each *epoch* (see §IV-B1) into several slots, and solves an optimization problem on the probability vector $\boldsymbol{\pi}$ similar to TASP, with the optimization objective of minimizing P_{total} (Eq. 7) at each slot. A traffic predictor is used that estimates the traffic peak at each slot.

Computing energy minimization: this workload distribution approach is similar with the total energy minimization in all respects except that it minimizes P_{comp} (instead of P_{total}).

The idea behind computing energy minimization approach is that reducing the computing energy will also reduce the cooling energy, and thus the total energy. The idea behind cooling energy minimization approach is that cooling energy depends on computing energy, therefore minimizing it would mean reducing the computing energy as well.

C. Simulation environment

Our study is based on a model of the ASU HPCI data center, which features 50 chassis of server blades (at the time of modeling). Figure 5 shows the layout of the assumed data center. We examine two constitution cases:

Homogeneous constitution: all chassis are of Dell 1855 (a=500, w=2020, $c = 0.04$/requests/sec).

Heterogeneous constitution: 20 chassis of Dell 1855 (parameters as in homogeneous) and 30 chassis of Dell 1955 (a=900, w=1590, $c = 0.01$/requests/sec).

The values of a and w are taken from experimental measurements done on blade server systems as published in [19]. The error of the linear model from the actual recordings is about 3%. The CoP curve used is given by [12]:

$$CoP(T) = 0.0068T^2 + 0.0008T + 0.48.$$

For input workload, we use a combination of the World Cup 1998 traces, for intensity, (http://ita.ee.lbl.gov/html/contrib/WorldCup.html) and the SPECweb, for workload generation. All calculations were done using MATLAB.

[2]In this context, "even balancing" means distributing the workload in such a way as to induce even utilization level across the servers, as opposed to evenly splitting the traffic.

TABLE I
TABLE OF SIMULATION FIGURES

Study	Data center constitution	Energy proportionality	Bar chart	Percentile savings
Server provisioning over no server provisioning	homogeneous	Case 1	Fig. 7	Fig. 8
		Case 2	Fig. 9	Fig. 10
	heterogeneous	Case 1	Fig. 11	Fig. 12
		Case 2	Fig. 13	Fig. 14
Workload management over load balancing under server provisioning	homogeneous	Case 1	Fig. 15	Fig. 16
		Case 2	Fig. 17	Fig. 18
	heterogeneous	Case 1	Fig. 19	Fig. 20
		Case 2	Fig. 21	Fig. 22

In this study, we variate the energy proportionality in terms of IPR, i.e.

$$P_{\text{idle}} = \text{IPR} \cdot P_{\text{peak}}, \tag{10}$$

along two IPR-LDR relation curves (Fig. 6):

Energy Proportionality Case 1: the IPR-LDR relation follows the right leg of the λ shape in Figure 4. We use the function $\text{LDR} = e^{-5\,\text{IPR}}$ (Fig. 6(a)) to *approximate* the relation, without claiming that the scatter plot necessarily follows this function. In this IPR-LDR relation, the power curves have a positive LDR, which means that they exhibit a large "mountain" followed by a smoother "valley". To emulate this case, we synthesize power curves using the following generic function:

$$p_i(u) = a_i u + w_i + b_i \frac{\sin(2\pi u)}{(u+1)^3 + 1},$$

where a and w are adjusted to match the desired IPR, and b is adjusted so that $P(u)$ matches the desired LDR. The power curves used are shown in Fig. 6(b).

Energy Proportionality Case 2: the IPR-LDR relation follows a straight line perpendicular to LDR=0 (Fig. 6(a)). This is a hypothetical progress toward true energy proportionality. It is used in contrast to the effects of the Case 1. The power curves used in this case follow the generic function $P(u) = a\,u + w$ and are shown in Fig. 6(c).

The study also considers three power density cases:

As-measured density (P_{peak} as measured): the power consumption range of the computing equipment is unaltered, i.e. it is $[\text{IPR}{\cdot}P_{\text{peak}}, P_{\text{peak}}]$.

Half density: (halved P_{peak}): the power consumption range of the computing equipment is halved, i.e. it is $[\text{IPR}{\cdot}P_{\text{peak}}/2, P_{\text{peak}}/2]$.

Double density: (doubled P_{peak}): the power consumption range of the computing equipment is doubled, i.e. it is $[2{\cdot}\text{IPR}{\cdot}P_{\text{peak}}, 2{\cdot}P_{\text{peak}}]$.

In the following graphs, each line corresponds to one combination of workload distribution, constitution and power density. Table I lists the graphs related to this study.

D. Effects of energy proportionality on the energy savings of server provisioning

This section compares the energy savings of server provisioning over no server provisioning under various energy proportionality and power density cases. It effectively addresses the question of *whether server provisioning will still provide energy savings under good energy proportionality of computing systems*.

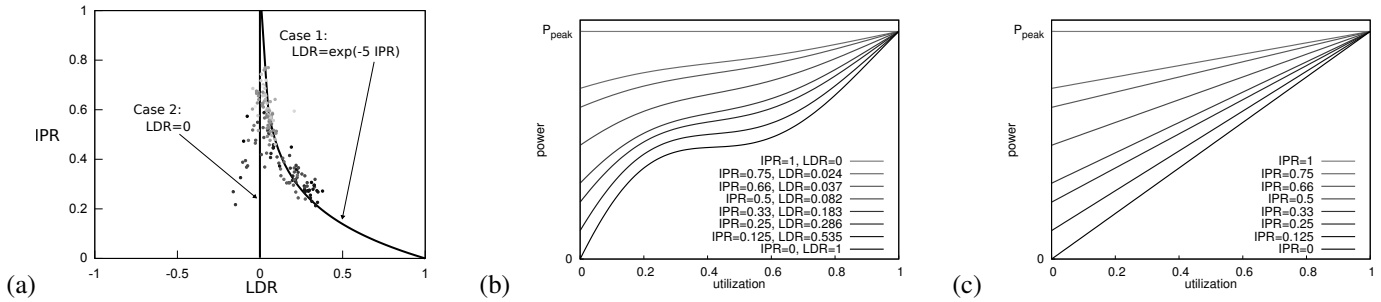


Fig. 6. (a) Two cases of IPR-LDR relation are examined: one that roughly follows the right leg of the λ shape in Fig. 4, and one that assumes an LDR of zero. (b) Synthesized power curves for case 1. (c) Synthesized power curves for case 2.

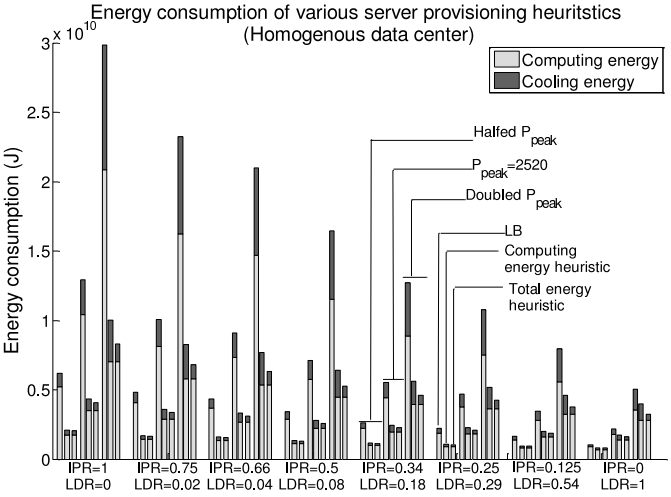


Fig. 7. Energy consumption of various server provisioning approaches for a homogeneous data center under Case 1.

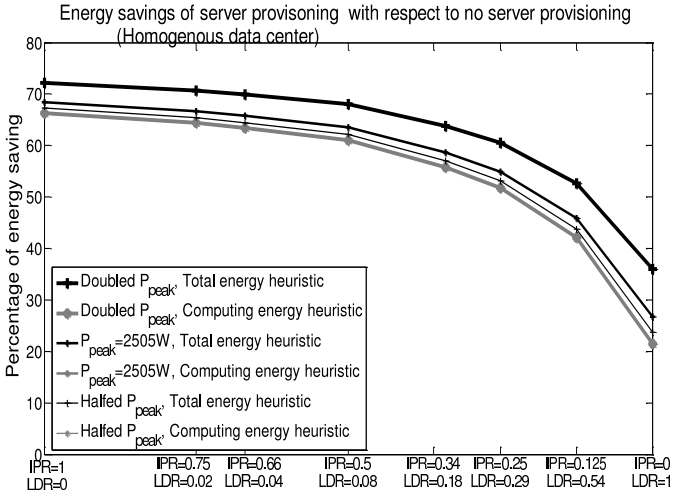


Fig. 8. Percentile savings over no server provisioning in Fig. 7.

Figures 7 and **8** show the energy savings of server provisioning over no server provisioning in a homogeneous data center, using load balancing as workload management, under Case 1. The figures show that power density has no effect to the percentile energy savings for the computing-energy heuristic (although it does affect the absolute value). In contrast, the total-energy heuristic's savings are favorably affected by density; this is because the difference in CoP of the thermal-aware server set choices is amplified by the increasing power density. Also, the savings are still significant even at IPR=0; this is an effect of the near-flat portion of the power curve around 30%-60%, thus consolidating workload to less servers without increasing the per-server power consumption. For IPR=0, the savings are greater than zero because of the workload consolidation over the flat portion (note the reduction in computing energy in Fig. 7), and due to cooling energy savings (note the reduction in cooling energy in the same figure).

Figures 9 and **10** show the energy savings of server provisioning over no server provisioning in a homogeneous data center, using load balancing as workload management, under Case 2. The figures show that the savings face larger reduction in this case: for IPR=LDR=0, the energy savings of the computing energy heuristic are the zero. The savings of the total-energy heuristic are significantly above zero because of the cooling-energy savings (note the difference in cooling energy for the total-energy heuristic in Fig. 9 for IPR=0).

Figures 11 and **12** show the energy savings of server provisioning over no server provisioning in a heterogeneous data center, using load balancing as workload management, under Case 1. The figures show that heterogeneity has preserving effect in energy savings. This is because server provisioning shuts off the less efficient servers and keeps the more efficient servers running (under no server provisioning, the less efficient servers would still be utilized).

Figures 13 and **14** show the energy savings of server provisioning over no server provisioning in a heterogeneous data center, using load balancing as workload management, under Case 2. The figures show that heterogeneity causes server provisioning to have greater-than-zero savings even ideal energy proportionality (IPR=LDR=0) for the computing-energy heuristic. This is because server provisioning takes advantage of the heterogeneity to chose more efficient servers over less efficient servers.

E. Effects of energy proportionality on the savings of workload distribution

This subsection compares the energy savings of various workload management approaches, over load balancing (LB), *under server provisioning*. It effectively addresses the question

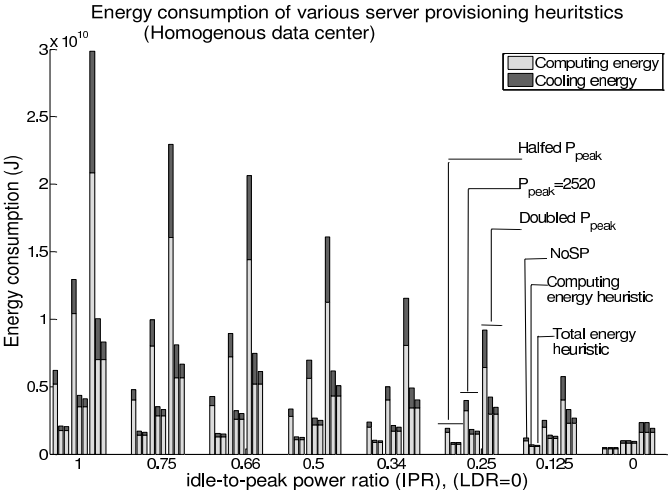


Fig. 9. Energy consumption of various server provisioning approaches for a homogeneous data center under Case 2.

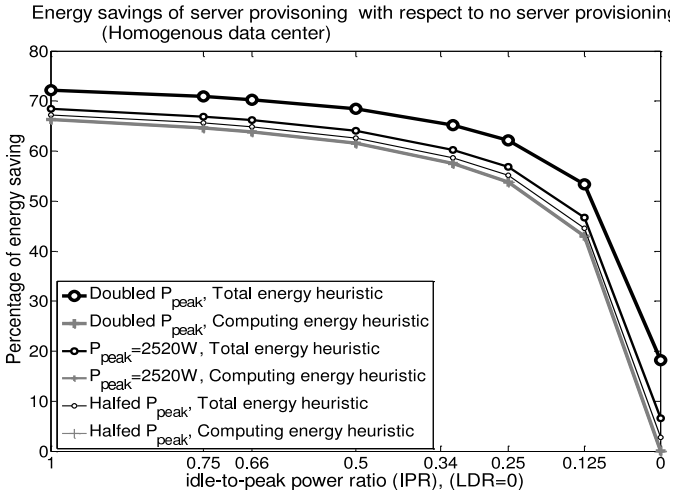


Fig. 10. Percentile savings over no server provisioning in Fig. 9.

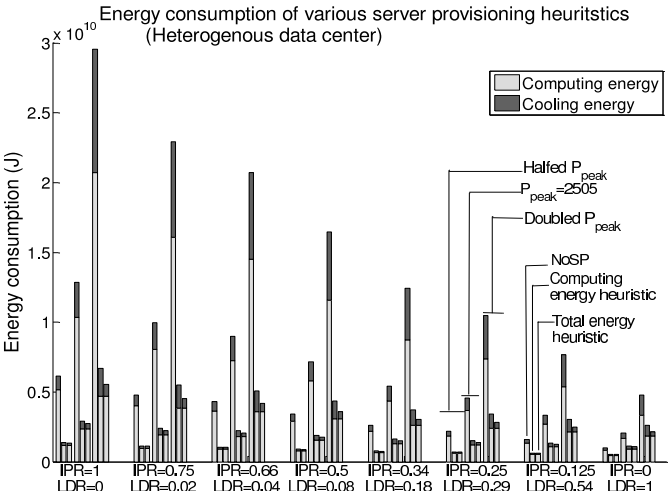


Fig. 11. Energy consumption of various server provisioning approaches for a heterogeneous data center under Case 1.

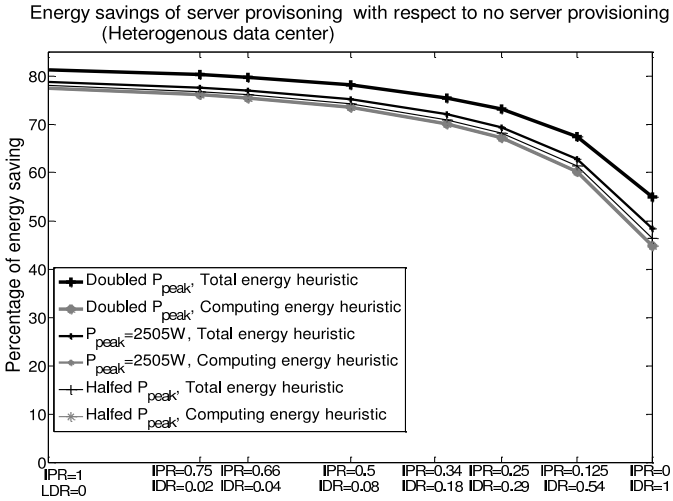


Fig. 12. Percentile savings over no server provisioning in Fig. 11.

of *whether load balancing under good energy proportionality renders "thermal-aware" solutions unnecessary*.

Figures 15 and **16** show the energy savings of the total-energy and computing-energy heuristics over the energy-oblivious load balancing, under thermal-aware server provisioning, for a homogeneous data center and for Case 1. The savings are increasing with increasing energy proportionality; this is because of the consolidation effect of the workload skewing over the "flat" portion of the power curve. This can be contrasted with **Figures 17** and **18** which show that the savings under a linear power consumption are almost negligible.

Figures 19 and **20** show the energy savings of the total-energy and computing-energy heuristics over the energy-oblivious load balancing, under thermal-aware server provisioning, for a heterogeneous data center and for Case 1. The savings are marginally better; this is because the active server set is mostly around 15-18 servers, which can easily fit in one type of servers, thus making it a homogeneous server set at most time. The slight difference is shown in **Figures 21** and **22**, where the savings of the workload heuristics are doubled in Case 2. The extra savings comes from thermally efficient selection of servers when the active server set size is large.

To demonstrate the fact that the server provisioning makes the active server set homogeneous for most of the time, we conducted a separate simulation using half the servers, thus forcing the active server set to be heterogeneous for most of the time. This resulted in energy savings that exceeded 50%, as shown in **Figure 23**.

V. Discussion

From the simulation above, it can be concluded that energy proportionality will necessarily diminish the savings of server provisioning. That would be true only in a homogeneous environment with no heat recirculation using a thermally oblivious scheme (Fig. 10. Server provisioning can be tuned toward energy-awareness to deliver significant savings.

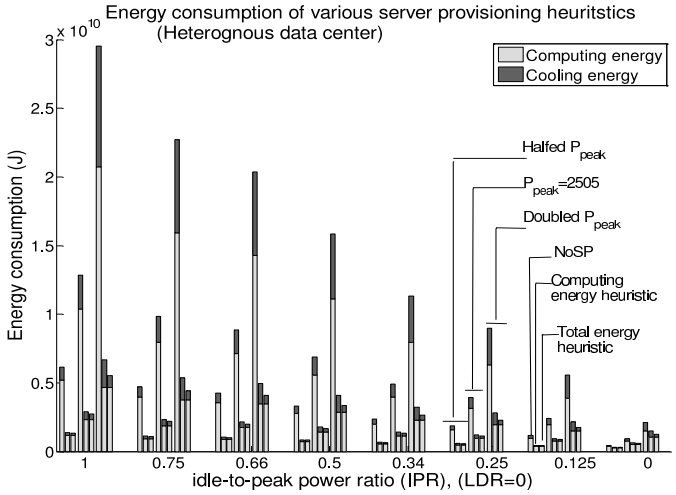


Fig. 13. Energy consumption of various server provisioning approaches for a heterogeneous data center under Case 2.

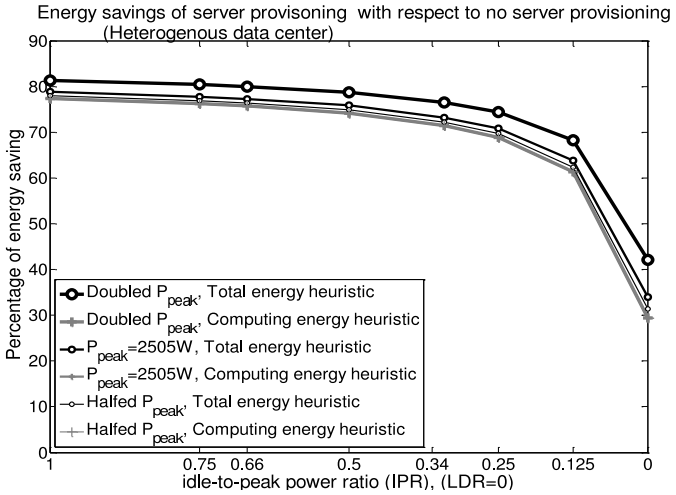


Fig. 14. Percentile savings over no server provisioning in Fig. 13.

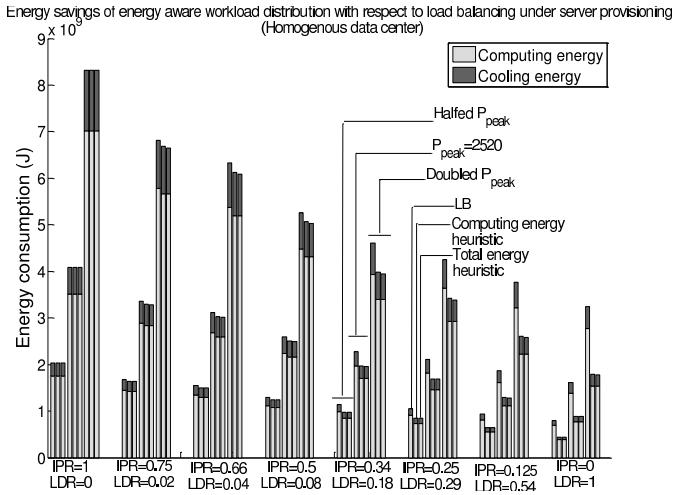


Fig. 15. Energy consumption of various workload management approaches for a homogeneous data center under Case 1.

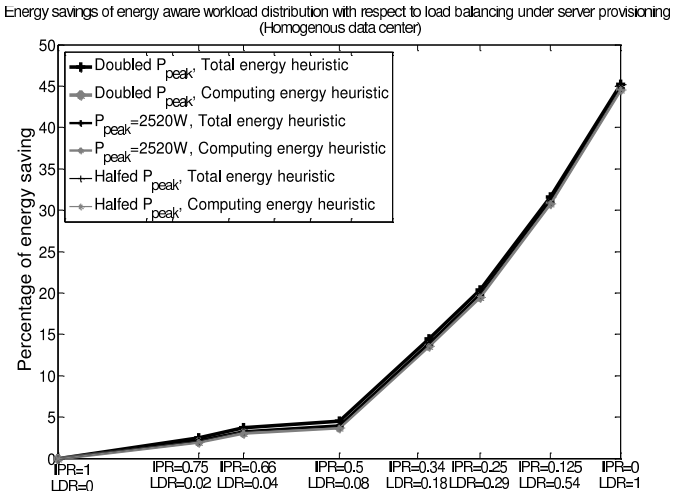


Fig. 16. Percentile savings over load balancing with server provisioning in Fig. 15.

Another significant observation is that non-linear energy proportionality can help in the savings of server provisioning. Provisioning can consolidate the workload to fewer servers, thus increasing the per-server utilization with minimal increase in energy consumption.

Lastly, heterogeneity in energy consumption plays a significant role in the energy savings over load balancing. This is because server provisioning and thermal-aware workload management will avoid giving workload to the less energy-efficient servers.

In general, the main benefit of energy proportionality is demonstrated in the difference between Fig. 24(a) and (b): a non-energy proportional does not have good energy efficiency at lower utilization levels, whereas a truly energy proportional has a *constant* energy efficiency. This property is what makes energy-proportional systems attractive; systems can be deployed and utilized at any level without considering any loses in efficiency (it will be one less parameter to consider during planning a data center).

However, the simulations in the previous section have revealed energy-saving benefits for LDR≠0. For example, for the power curve in Fig. 24(c), with LDR>0, is shown to yield energy savings when server provisioning is used. This is because the server provisioning consolidates the workload to fewer servers, thus increasing the per-system utilization to a more energy-efficient level, for example from 30% to 60%.

On the other hand, a system with LDR<0, e.g. with a power curve in Fig. 24(d), will have increasing energy-efficiency as the utilization approaches zero. Such a system would be most suitable for applications where the utilization is close to zero.

VI. Concluding Remarks

The above observations shift the balance of savings and warrant a more careful study. Similarly, the study in §III-A should be repeated using *larger* and especially *more representative* pools of data. However, it is clear that energy-proportional computing will soon become a reality, and its

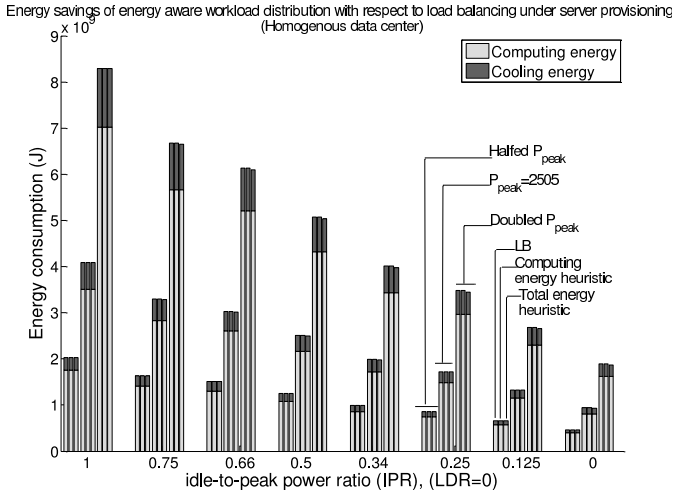


Fig. 17. Energy consumption of various workload management approaches for a homogeneous data center under Case 2.

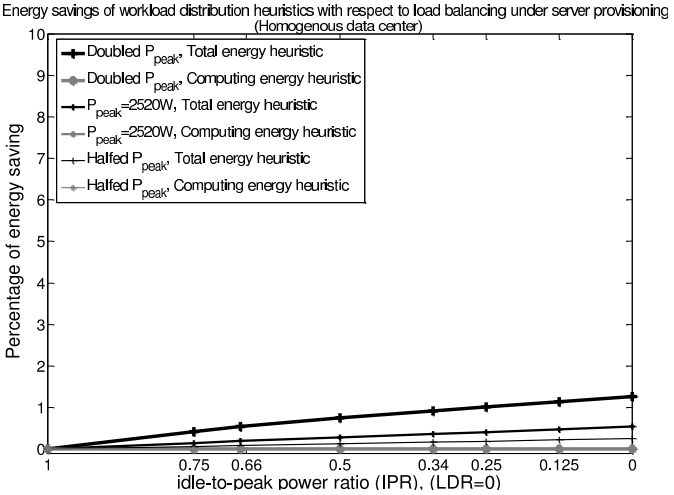


Fig. 18. Percentile savings over load balancing with server provisioning in Fig. 17.

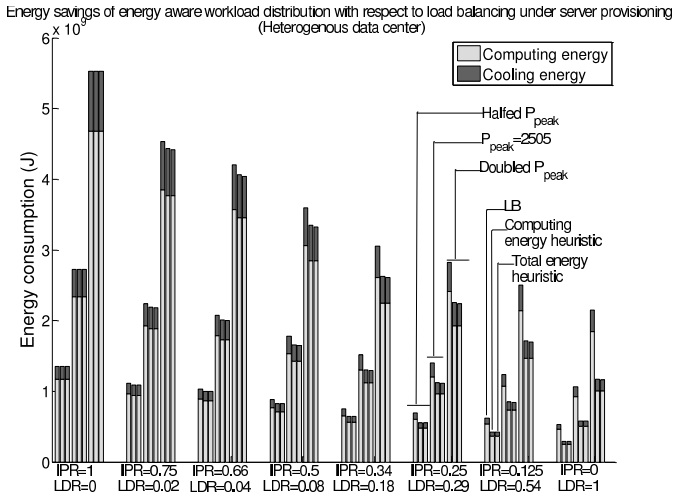


Fig. 19. Energy consumption of various workload management approaches for a heterogeneous data center under Case 1.

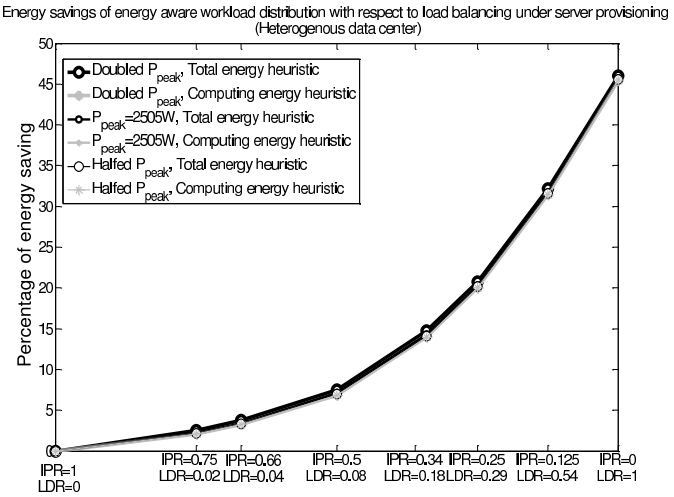


Fig. 20. Percentile savings over load balancing with server provisioning in Fig. 19.

effects on the design and management of data centers needs to be re-examined.

Acknowledgments

We would like to thank Luiz Barroso at Google Inc., and Partha Ranganathan at Hewlett-Packard, for their insightful comments.

References

[1] D. Kusic, J. O. Kephart, J. E. Hanson, N. Kandasamy, and G. Jiang, "Power and performance management of virtualized computing environments via lookahead control," *Cluster Computing*, vol. Volume 12, pp. 1–15, 2009.

[2] P. Padala, K.-Y. Hou, K. G. Shin, X. Zhu, M. Uysal, Z. Wang, S. Singhal, and A. Merchant, "Automated control of multiple virtualized resources," in *Proc. of the Europe Conference on Comp. Sys.*, March 2009.

[3] G. Chen, W. He, J. Liu, S. Nath, L. Rigas, L. Xiao, , and F. Zhao, "Energy-aware server provisioning and load dispatching for connection-intensive internet services," in *NSDI'08: Proceedings of the 5th USENIX Symposium on Networked Systems Design and Implementation*. Berkeley, CA, USA: USENIX Association, 2008, pp. 337–350.

[4] J. Chase, D. Anderson, P. Thakar, A. Vahdat, and R. Doyle, "Managing energy and server resources in hosting centers," in *SOSP '01: Proceedings of the eighteenth ACM symposium on Operating systems principles*. New York, NY, USA: ACM, 2001, pp. 103–116.

[5] L. A. Barroso and U. Hölzle, "The case for energy-proportional computing," *Computer*, vol. 40, no. 12, pp. 33–37, 2007.

[6] ASHRAE, "Datacom equipment power trends and cooling applications," Atlanta, GA, 2005.

[7] P. Bohrer, E. N. Elnozahy, T. Keller, M. Kistler, C. Lefurgy, C. McDowell, and R. Rajamony, "The case for power management in web servers," pp. 261–289, 2002.

[8] P. Barford and M. Crovella, "Generating representative web workloads for network and server performance evaluation," *SIGMETRICS Perform. Eval. Rev.*, vol. 26, no. 1, pp. 151–160, 1998.

[9] Y. Chen, A. Das, W. Qin, A. Sivasubramaniam, Q. Wang, , and N. Gautam, "Managing server energy and operational costs in hosting centers," *SIGMETRICS Performance Evaluation Review*, vol. 33, no. 1, pp. 303–314, 2005.

[10] M. Elnozahy, M. Kistler, and R. Rajamony, "Energy conservation policies for web servers," in *USITS'03: Proceedings of the 4th conference on USENIX Symposium on Internet Technologies and Systems*. Berkeley, CA, USA: USENIX Association, 2003, pp. 8–8.

[11] P. Ranganathan, P. Leech, D. Irwin, and J. Chase, "Ensemble-level power

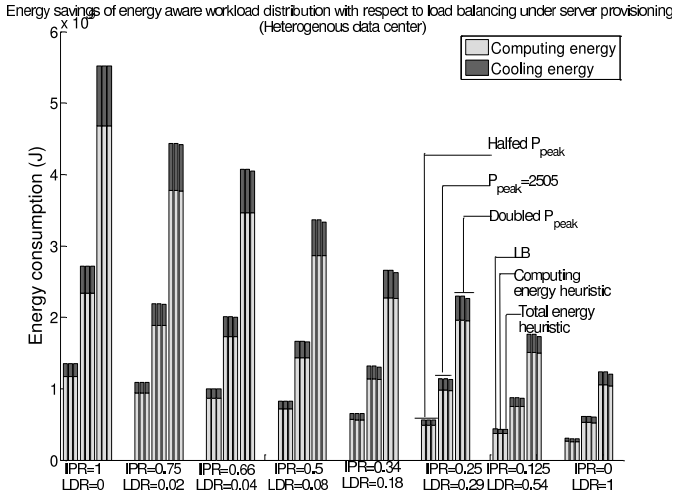


Fig. 21. Energy consumption of various workload management approaches for a heterogeneous data center under Case 2.

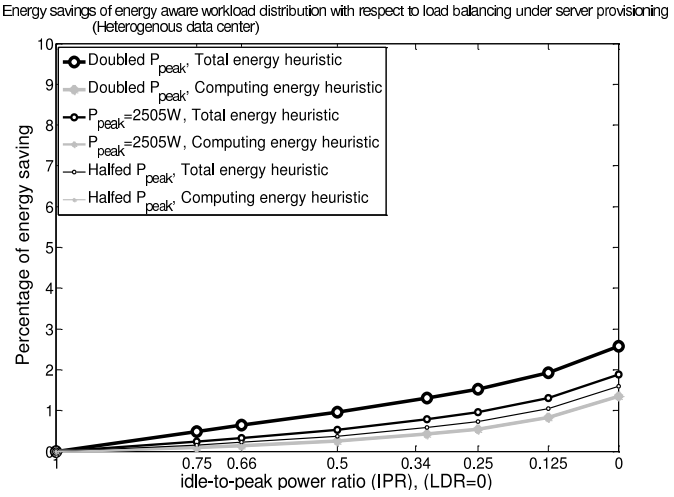


Fig. 22. Percentile savings over load balancing with server provisioning in Fig. 21.

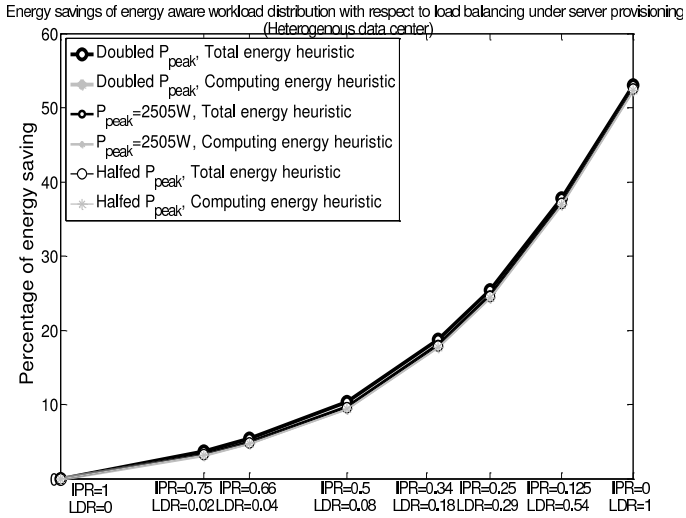


Fig. 23. Percentile savings over load balancing with server provisioning in a half-sized data center.

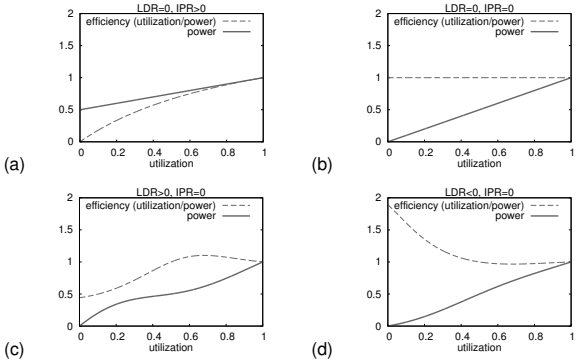


Fig. 24. Energy efficiency for various energy proportionality cases: (a) non-proportional computing (IPR>0); (b) ideal energy-proportional computing; (c) near-proportional with LDR>0; (d) near-proportional with LDR<0.

management for dense blade servers," in *Computer Architecture, 2006. ISCA '06. 33rd International Symposium on*, 0-0 2006, pp. 66–77.

[12] J. Moore, J. Chase, P. Ranganathan, and R. Sharma, "Making scheduling "cool": temperature-aware workload placement in data centers," in *ATEC '05: Proceedings of the annual conference on USENIX Annual Technical Conference*. Berkeley, CA, USA: USENIX Association, 2005, pp. 5–5.

[13] Q. Tang, S. K. S. Gupta, and G. Varsamopoulos, "Energy-efficient thermal-aware task scheduling for homogeneous high-performance computing data centers: A cyber-physical approach," *IEEE Trans. Parallel Distrib. Syst.*, vol. 19, no. 11, pp. 1458–1472, 2008.

[14] T. Mukherjee, A. Banerjee, G. Varsamopoulos, S. K. S. Gupta, and S. Rungta, "Spatio-temporal thermal-aware job scheduling to minimize energy consumption in virtualized heterogeneous data centers," *Computer Networks*, June 2009. [Online]. Available: http://dx.doi.org/10.1016/j.comnet.2009.06.008

[15] L. Wang, A. J. Younge, T. R. Furlani, G. von Laszewski, J. Dayal, and X. He, "Towards thermal aware workload scheduling in a data center," in *Proceedings of the 10th International Symposium on Pervasive Systems, Algorithms and Networks*, 2009, pp. 116–122.

[16] L. Parolini, N. Toliaz, B. Sinopoli, and B. H. Krogh, "A cyber-physical systems approach to energy management in data centers," in *ACM ICCPS '10*, April 2010.

[17] R. Das, J. O. Kephart, J. Lenchner, and H. Hamann, "Utility-function-driven energy-efficient cooling in data centers," in *IEEE/ACM International Conference on Autonomic Computing, ICAC10*, june 2010, pp. 61–70.

[18] K. W. Cameron, "The challenges of energy-proportional computing," *Computer*, vol. 43, pp. 82–83, 2010.

[19] T. Mukherjee, G. Varsamopoulos, S. K. S. Gupta, and S. Rungta, "Measurement-based power profiling of data center equipment," in *IEEE International Conference on Cluster Computing.*, Sept 2007, pp. 476–477.

[20] Z. Abbasi, G. Varsamopoulos, and S. K. S. Gupta., "Thermal-aware server provisioning and workload management for internet data centers," in *HPDC 2010: Proceedings of the ACM International Symposium on High Performance Distributed Computing*, Jun. 2010.