Non-IT Energy Accounting in Virtualized Datacenter

Weixiang Jiang[1], Shaolei Ren[2], Fangming Liu*[1], and Hai Jin[1]
[1]Key Laboratory of Services Computing Technology and System, Ministry of Education,
School of Computer Science and Technology, Huazhong University of Science and Technology, China
[2]University of California, Riverside, USA

***Abstract*—Energy accounting plays a crucial role in datacenter energy management, wherein the energy consumption of non-IT units (e.g., UPS and cooling system) makes up a significant portion. However, it is challenging to *fairly* account for non-IT energy on an individual VM basis, because the non-IT units are shared by multiple VMs in a virtualized datacenter and only the system-level non-IT energy consumption can be measured. Existing policies, e.g., equally or proportionally allocating non-IT energy to VMs based on their IT energy, are not fair, in the sense that they can not satisfy a set of desired axiomatic principles of fair allocation. In this paper, we propose LEAPS, a Lightweight Energy Accounting Policy based on a provably fair methodology called Shapley value. We evaluate it using real-world datacenter trace and demonstrate that, compared to original Shapley value approach that has an exponential complexity, LEAPS yields almost the same energy accounting result within a maximum relative error less than 6.97%, while having a negligible computation time.**

I. Introduction

Nowadays, more and more companies migrate their businesses to cloud, leading to explosive expansion of datacenter scale and rapidly increasing energy consumption. Hence, they are not only accountable for the energy usage of their own facilities, but also the energy usage in public clouds and third-party datacenters, which have been mandated by government regulations and pressured by organizations such as Greenpeace [1]. For example, Apple[1] and Akamai have announced to include energy usage in cloud and third-party datacenters as part of their electricity footprint [2], [3]. In a datacenter, the cloud tenants not only consume the energy of IT units (e.g., servers and switches), but also the energy of non-IT units (e.g., UPS and cooling system). Efforts have been paid to provide energy accounting for IT units at different levels (e.g., server, VM, application and even component [4], [5], [6], [7]). However, the non-IT energy of individual tenant still remains ambiguous.

Obviously, as each tenant owns several VMs, the first and also crucial step is to measure non-IT energy consumption on an individual VM basis. The non-IT energy in a datacenter mainly consists of energy consumption of cooling system, and energy loss from uninterruptible power supply (UPS) and power distribution unit (PDU) [8]. Till now, these non-IT units still consume substantial energy in a datacenter. During $2011 \sim 2014$, the world-wide average PUE of datacenters only reduced from 1.89 to 1.7 [9]. Because of geographic and climate conditions, energy-intensive mechanical cooling is the only option for many datacenters, making non-IT energy a significant fraction of the total datacenter-level energy. Even liquid cooling only reduces $21 \sim 22\%$ cooling energy as shown by two mainstream liquid cooling companies [10], [11]. Besides, the voltage conversion efficiency of UPS in today's datacenters is limited to $80 \sim 95\%$, leading to $5 \sim 20\%$ energy loss [12], [13]. In an average datacenter where centralized UPS and chillers are commonly used, the non-IT part can take up 30% or even 50% of the total energy [14], [15].

Nonetheless, the non-IT energy accounting is difficult, as the non-IT units are shared by multiple VMs and inherently non-divisible. An empirical method is to attribute the non-IT energy to different VMs in proportion to their IT energy usage. This energy accounting policy, albeit simple, is not fair. This is because the non-IT energy grows non-linearly (as shown later in Sec. II), which implies the same IT energy increment may contribute different non-IT energy. Besides, non-IT unit (e.g., UPS) also has a static energy[2] when active [13], [16]. We may equally split the static energy to each VMs or attribute it to VMs in proportion to their IT energy. *But which one is better or fairer? Further, is there a convincing solution to the above non-IT energy accounting problem that has a certain theoretical basis?*

Intuitively, we find the non-IT energy accounting problem is similar to the revenue allocation problem in game theory: How do we determine each player's contribution in a cooperative game and what payoff should each individual player receive? In our context, the non-IT energy of the whole datacenter represents the total "revenue" and the VMs are the players. To solve the revenue allocation problem, Shapley value was proposed, which has been proven to be the only fair method

*The corresponding author is Fangming Liu (fmliu@hust.edu.cn). This work was supported in part by the National Key Research & Development (R&D) Plan under grant 2017YFB1001703, in part by NSFC under Grant 61722206 and 61761136014 and 392046569 (NSFC-DFG) and 61520106005, in part by the National 973 Basic Research Program under Grant 2014CB347800, in part by the Fundamental Research Funds for the Central Universities under Grant 2017KFKJXX009, and in part by the National Program for Support of Top-notch Young Professionals in National Program for Special Support of Eminent Professionals.

[1]Apple not only owns and operates datacenters, but also rents servers in colocation datacenters and deploys on-demand service in cloud for additional capacity [2].

[2]The static energy is to keep a device active when it is idle. For example, UPS still consumes energy even when there is no load on it.

for revenue sharing (in the sense of satisfying a set of desired axiomatic principles listed later in Sec. IV-B) [17], [18]. Nonetheless, applying it in our context has two major challenges. First, calculation of Shapley value requires the knowledge of non-IT energy consumption for every possible subset of VMs. In practice, however, we can only measure the energy consumption for each non-IT unit as a whole when all or a certain set of VMs are active. Second, and more critically, deriving Shapley value has an exponential complexity that quickly becomes computationally prohibitive even for 20+ VMs, but a real-world datacenter often has several hundred or even thousand VMs.

In this paper, we address the above challenges of using Shapley value for fair non-IT energy accounting in virtualized datacenters. Specifically, we approximate the non-IT energy consumption as a quadratic function of the VMs' IT energy, which allows us to account for VMs' non-IT energy in real time and also has a high accuracy (compared to the original Shapley value calculation). We call our method LEAPS (Lightweight Energy Accounting Policy based on Shapley value). LEAPS is derived from Shapley value and easy to implement with an interesting insight: it attributes dynamic energy of non-IT systems to tenants in proportion to their IT energy usage, and equally splits the static energy of non-IT systems among all active VMs.

Specifically, we make the following major contributions in this paper:

- We demonstrate that the non-IT energy accounting problem should be formulated as a cooperative game and Shapley value provides the theoretical ground truth with four axiomatic principles that guarantee the fairness.
- We study the energy consumption characteristics of different non-IT units in current datacenters and propose LEAPS, which leverages a quadratic function to approximate energy usage of each non-IT unit. LEAPS is derived from Shapley value, but reduces the complexity of original Shapley value from $O(2^N)$ to $O(N)$.
- We evaluate the accuracy of LEAPS using real-world datacenter power traces. The results demonstrate, compared to original Shapley value approach that has an exponential complexity, LEAPS yields almost the same energy accounting result within maximum relative error of less than 6.97%, and a negligible computation time.

II. Energy Consumption Characteristic of Non-IT Units

In this section, we first provide some backgrounds to the characteristics of non-IT units' energy consumption in a datacenter by measurements and survey.

A. *Measurement Platform*

Our measurements are conducted in a real-world datacenter as shown in Fig. 1. The datacenter accesses the grid through a transformer station and delivers the electricity to UPS and cooling system, respectively. The UPS transfers the high voltage alternating current (AC) to the low voltage direct current (DC) and accesses the battery as the backup power, then it makes a reverse transformation to supply the IT units with low voltage AC [19]. The IT units consists of 12 cabinets that house 346 computing nodes, 16 disk arrays, 10 GPU nodes and network switches, with a peak-rated power of ∼120kW. In addition, our datacenter is equipped with 7 power distribution management modules (PDMM) to monitor the power[3] of each server cabinet for circuit overload protection, which can provide us the IT units' power data (i.e., UPS power output) through an RS485 field-bus. Due to the datacenter locates in the center of the city, it adopts two precision air conditioner to guarantee the cooling performance. For the UPS power input and cooling system's power, we use a Fluke 1738 three-phase power logger [20] to record them. The UPS's power loss can be calculated by the difference of the power consumption recorded by PDMM and Fluke Power Logger.

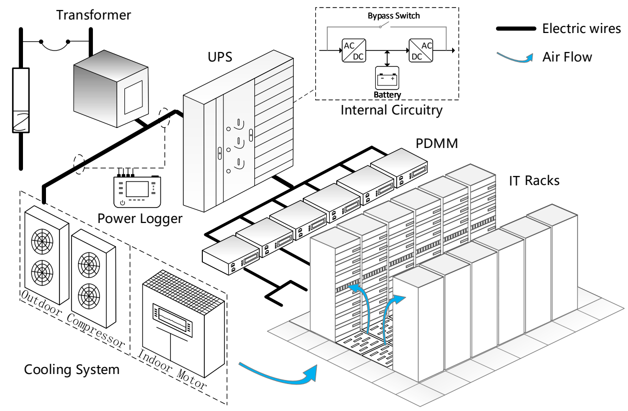


Fig. 1. Power architecture of the datacenter.

B. *Power Distribution Efficiency*

To lower the voltage and access the battery, the UPS will make AC/DC conversions during electricity delivering. However, such conversions can lead to $5\% \sim 20\%$ power loss [12], [13]. An earlier study of large UPS systems from Schneider [13] reports the UPS loss grows quadratically with the IT power load (i.e., $F(x) = 0.03455x^2 + 0.00959x + 0.03234$, where $F(x)$ is power loss and x is IT power load). Accordingly, we make a measurement to investigate the correlation between the UPS power loss and its power load in our datacenter.

Fig. 2 shows the measurement results of the UPS power loss. In particular, the UPS power loss can be approximated by the least square method:

$$F(x) = 0.0003x^2 + 0.0205x + 2.8628, \tag{1}$$

where x is the IT power load. The above observations are similar to Schneider's report [13]: the quadratic term is due to the UPS's circuit heat that increases quadratically with current

[3]Power usually measures the energy consumed per second. Energy can span a long timescale and is equivalent to power when the accounting period is one second.

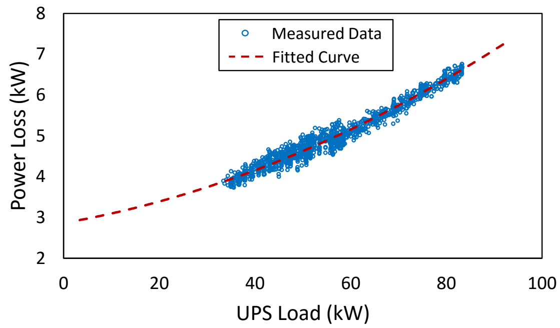


Fig. 2. Power loss of UPS.

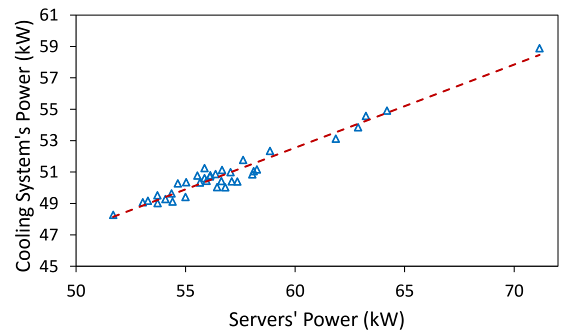


Fig. 3. Cooling system's power at the outside air temperature of 30°C.

(sometimes referred to as "I-squared R" losses), while the static term represents idle power to keep UPS active.

Due to "I-squared-R" losses, PDUs also incur an energy loss proportional to the square of the IT power load [15]. The energy losses finally dissipate in the form of heat.

C. *Cooling Power Model*

We next investigate the power correlation between cooling system and IT units. There are multiple cooling approaches in datacenters. At present, the precision air conditioning refrigeration, liquid cooling and outside air cooling (OAC) are the most popular cooling methods in current datacenters [21].

Precision air conditioner refrigeration: Fig. 3 summarizes the power measurement results of cooling system and IT units in our datacenter, which are collected during about one and a half months when the outside temperature is about 30°C. The power correlation between the precision air conditioner refrigeration and IT units can be approximated by a linear function:

$$F(x) = 0.53x + 20.749, \tag{2}$$

with $R^2 = 0.9528$, where is x is the IT power load. Note that the IT power load in a datacenter typically stays in a certain utilization range (as shown later in Fig. 6) instead of varying between zero and the maximum. Thus, there is no need to approximate the cooling power for the entire range of IT power loads. Usually the energy efficiency ratio (EER), i.e., the heat that can be transported by an air conditioner per watt, is fixed. The heat dissipated by IT units is roughly equal to their power consumption. So to keep the indoor temperature constant, the power cost of air conditioner would grow linearly with IT power load.

Liquid cooling uses chilled water rather than cold air to dissipate heat from server components. The liquid heated by servers exchanges heat with cold facility water from outside cooling tower, and then flows back to chill servers again. Particularly, its power consumption is quadratic in the IT power load as reported by [15] (i.e., $F(x) = 742.8x^2 + 1844.6x + 538.7$, where x is its IT power load).

Outside air cooling uses the cold air from the outside to chill the servers directly. This method is unusual because it requires that datacenters must locate in cold areas, such as seaside, and needs air filtration system to protect electron device from dust and corrosion. Its power usage is mainly caused by blowers. The cooling efficiency of this method highly depends on the temperature difference between outside air and server components. As reported by [14], the power of outside air cooling can be express as $F(x) = kx^3$, where x is IT power and $k \geq 0$ is related to the outside temperature.

Overall, a datacenter may have multiple non-IT units with different power consumption characteristics. In this section, we summarize the power consumption characteristics of the popular non-IT units in current datacenters, including precision air conditioner, UPS, chilled water and outside air cooling system, whose power consumption can be formulated as linear, quadratic and cubic function, respectively.

III. Non-IT Energy Accounting Definition and Current Policies

In this section, we first describe the problem of non-IT energy accounting. Then we overview the existing energy accounting policies. For ease of reference, Table I lists the notations used in the rest sections.

A. *Problem Statement*

Datacenters usually manage and provide their compute capacity to tenants in the form of VMs [22]. So suppose that there are N running VMs and M different non-IT units in a datacenter. The entire set of VMs is denoted by $\mathcal{N}$. The non-IT units under consideration include UPS and cooling system, etc. Each non-IT unit serves multiple VMs, while each VM affects energy consumption of multiple non-IT units.

We denote $\mathcal{M}_i$ as the set of non-IT units whose energy is affected by VM i. Letting $\mathcal{N}_j \subseteq \mathcal{N}$ be the subset of VMs that affects non-IT unit j, energy consumption of j can be written as $P_j = F_j(\sum_{i\in\mathcal{N}_j} P_i)$, where P_i is IT energy consumption of VM i and $F_j(\cdot)$ is referred to as the energy function that relates VMs' IT energy consumption to that of non-IT unit j. Denote $\Phi_{ij} \geq 0$ as VM i's share from non-IT unit j's energy, and thus we have $P_j = \sum_{i\in\mathcal{N}_j} \Phi_{ij}$.

Definition 1: Formally, the non-IT energy accounting problem is to *fairly* determine each VM i's non-IT energy $\Phi_i = \sum_{j\in\mathcal{M}_i} \Phi_{ij}$.

TABLE I
NOTATIONS

Symbol	Description
N	Number of VMs
M	Number of non-IT units
$\mathcal{N}_j$	The set of VMs that affect non-IT unit j
$\mathcal{M}_i$	The set of non-IT units whose energy is affected by VM i
$F_j(\cdot)$	The energy function that relates VMs' energy consumption to that of non-IT unit j
Φ_{ij}	VM i's non-IT energy share from non-IT unit j
$\mathcal{X}$	The subset of $\mathcal{N}_j$
P_j	Energy consumption of non-IT unit j
P_i	Energy consumption of VM i
$P_{\mathcal{X}}$	Aggregated energy consumption of the VMs in the set $\mathcal{X}$
n_j	The size/cardinality of set $\mathcal{N}_j$
$r_{\mathcal{X}}$	The size/cardinality of set $\mathcal{X}$
δ_x	The deviation of a quadratic non-IT power approximation when IT power is x
a_j, b_j, c_j	The coefficients of an approximated quadratic function for non-IT unit j

However, it is non-trivial to *fairly* decompose $F_j(\cdot)$ into multiple shares for each VM, because one can only measure the energy consumption P_j of a shared non-IT unit as a whole. Further, non-IT unit j's energy function $F_j(\cdot)$ is non-linear in VMs' energy (e.g., UPS energy loss grows quadratically in its load and outside air economizer energy grows cubically).

B. *Existing Non-IT energy Accounting Policies*

We now introduce the existing energy accounting policies.

Policy 1: $\Phi_{ij} = F_j/|\mathcal{N}_j|$: This policy states that the non-IT energy consumption of each VM is equal to the total energy consumption divided by the number of VMs. That is, each VM gets an equal share of the total non-IT energy consumption.

Policy 2: $\Phi_{ij} = F_j \cdot P_{i\in\mathcal{N}_j}/\sum_{l\in\mathcal{N}_j} P_l$. This policy states that the total non-IT energy is attributed in proportion to each VM's average IT energy over a predefined accounting period (e.g., a second or an hour). This policy is commonly used for charging tenants' non-IT energy consumption in co-location datacenters [23].

Policy 3: $\Phi_{ij} = F_j(P_i + P_{\mathcal{X}}) - F_j(P_{\mathcal{X}})$, where $P_{\mathcal{X}} = \sum_{k\in\mathcal{N}_j\setminus\{i\}} P_k$. This policy states that the energy attributed to a VM equals to its marginal non-IT energy contribution, i.e., the energy variation of a non-IT unit when a VM starts to run, while all other VMs' power remains unchanged.

Each of the above policies seems reasonable but empirical. Next we will introduce Shapley value as the ground truth for non-IT energy accounting and reveal the fundamental drawbacks of existing policies analytically (Sec. IV-C) and quantitatively (Sec. VII).

IV. SHAPLEY VALUE AS GROUND TRUTH

We now introduce a unique revenue allocation rule called Shapley value [17], [18] and its four axioms for revenue allocation. Then we demonstrate how existing policies violate one or more of the four axioms.

A. *Shapley Value*

Fairly sharing the total revenue among multiple players has been studied extensively in the cooperative game. There are two important elements in the revenue sharing problem: (*i*) $\mathcal{N} = \{1, 2, ..., n\}$ representing the set of n players, i.e., n VMs in our context, and (*ii*) $v(\mathcal{X})$ is a characteristic function, describing the revenue that a subset of players $\mathcal{X} \subseteq \mathcal{N}$ can generate in the game, i.e., the non-IT energy consumption generated by a subset of VMs in our context, which can be denoted as $F_j(P_{\mathcal{X}})$, $P_{\mathcal{X}} = \sum_{k\in\mathcal{X}} P_k$. Applying Shapley value in our context, the energy share of non-IT unit j attributed to VM i is calculated as

$$\Phi_{ij} = \sum_{\mathcal{X}\subseteq\mathcal{N}_j\setminus\{i\}} \frac{|\mathcal{X}|!(|\mathcal{N}_j| - |\mathcal{X}| - 1)!}{|\mathcal{N}_j|!} \cdot [F_j(P_{\mathcal{X}} + P_i) - F_j(P_{\mathcal{X}})] \tag{3}$$

where $\mathcal{X}$ is a subset of VMs (excluding VM i) in $\mathcal{N}_j$ supported by non-IT unit j, and $P_{\mathcal{X}} = \sum_{k\in\mathcal{X}} P_k$. Obviously, there are 2^N subsets $\mathcal{X} \subseteq \mathcal{N}$, resulting a total complexity of $O(2^N)$ for calculating Φ_{ij}.

We explain the implication of Shapley value in (3) as follows. Suppose that VMs join the non-IT unit sequentially, and consider a certain subset $\mathcal{X}$ of VMs that has already joined the non-IT unit j before VM i. Then, $F_j(P_{\mathcal{X}} + P_i) - F_j(P_{\mathcal{X}})$ is the marginal contribution of VM i to the non-IT unit j's energy increase. Note that the subset $\mathcal{X}$ of VMs can join the system in $|\mathcal{X}|!$ ways due to all different permutations, while the VMs that join the non-IT unit after VM i can happen in $(|\mathcal{N}_j| - |\mathcal{X}| - 1)!$ ways. The term $|\mathcal{N}_j|!$ in the denominator is to take the average of all the possible permutations of VMs joining the non-IT unit. Thus, by taking the average, the non-IT energy share of VM i is obtained.

B. *Why Shapley Value Acts as Ground Truth?*

The non-IT energy accounting problem can be essentially viewed as attributing a shared cost/payoff to individual players, which is a classic problem in cooperative game theory with fairness as a key consideration [17], [18]. While there is no uniform definition for fairness, prior studies have commonly used four axiomatic principles, and an allocation policy (i.e., energy accounting policy) satisfying all of them is said to be fair [24], [25], [26]. Below, we introduce these four axioms and explain them in our context:

Efficiency. The sum of accounted non-IT energy by individual VMs should be equal to the total non-IT energy, i.e., $\sum_{i\in\mathcal{N}_j} \Phi_{ij} = P_j$.

Symmetry. If two VMs are interchangeable and indistinguishable for their contribution to non-IT energy increase, they should account for the same non-IT energy, i.e., if $F_j(\sum_{i\in\mathcal{X}\cup\{k\}} P_i) = F_j(\sum_{i\in\mathcal{X}\cup\{l\}} P_i)$ for any $\mathcal{X} \subseteq \mathcal{N}_j \setminus \{k, l\}$, then $\Phi_{k,j} = \Phi_{l,j}$.

TABLE II
EXAMPLE OF THREE VMS' IT ENERGY USAGES IN DIFFERENT TIME INTERVALS (KW·S).

VM \ IT Energy	t_1	t_2	t_3	$T = t_1 + t_2 + t_3$
#1	28	22	13	63
#2	24	9.2	10	43.2
#3	10	14.2	19	43.2

TABLE III
HOW EXISTING ENERGY ACCOUNTING POLICIES VIOLATE THE AXIOMS THAT DEFINE THE FAIRNESS

Policy	Efficiency	Symmetry	Null Player	Additivity
I			×	
II		×		×
III	×	×		

Null player. That is, if the non-IT energy does not change when we add or remove a VM, zero non-IT energy should be attributed to this VM (also called null player in a game), i.e., if $F_j(\sum_{l\in\mathcal{X}\cup\{i\}} P_l) = F_j(\sum_{l\in\mathcal{X}} P_l)$ for any $\mathcal{X} \subset \mathcal{N}_j$, then $\Phi_{i,j} = 0$.

Additivity. If two coalition games are combined, then for any players, the sum of the gains in the two coalition games should be equal to the gain in the combined one. That is, if we break the energy accounting interval T into multiple time intervals $[t_1, t_2, ..., t_n]$, then T is the combined game of individual game $[t_1, t_2, ..., t_n]$. *Additivity* requires the sum of non-IT energy attributed to a VM in each time interval should be equal to the non-IT energy attributed to the VM in time T, i.e., $\sum_{t=t_1}^{t_n} \Phi_{i,j,t} = \Phi_{i,j,T}$, for any $T = t_1 + t_2 + ...t_n$.

Note that the four axioms are the principles for non-IT energy accounting rather than assumptions. The Shapley value is proven to be the **only** allocation rule that satisfies all of the four axiomatic principles [17], [18]. Hence, we argue that the Shapley value provides a naturally fair and intuitive solution for non-IT energy accounting.

C. Existing Policies against Axioms

Now we show the drawbacks of existing policies against the four axioms.

Policy 1 violates *Null player*. It disregards the differences between VMs, and VMs always get positive non-IT energy share according to it.

Policy 2 violates *Symmetry* and *Additivity*. We consider the UPS in our datacenter whose power loss can be expressed as $F(x) = 0.0003 \cdot x^2 + 0.0205 \cdot x + 2.8628$, where x is the power load of IT unit. To clearly show the *Symmetry* and *Additivity* violation, we consider three VMs and an energy accounting time period $[t_1, t_2, t_3]$ (each representing 1 second). Table II shows an example of the average IT energy usages of VM #1, VM #2 and VM #3 in each second. The total energy loss of UPS calculated by $F(x)$ during $[t_1, t_2, t_3]$ is 13.95 kW·s. According to Policy 2, the three VMs get a UPS loss share of 5.84 kW·s, 3.95 kW·s and 4.16 kW·s during $[t_1, t_2, t_3]$. However, if we consider another energy accounting interval of $T = t_1 + t_2 + t_3$, we can see that the three VMs have an energy usage of 63 kW·s, 43.2 kW·s and 43.2 kW·s during time T as shown in Table II. It results in the satisfaction of *Symmetry* condition for VM #2 and #3 during T, and the three VM get a UPS loss share of 5.884 kW·s, 4.03 kW·s and 4.03 kW·s. We can see that Policy 2 produces different results with different energy accounting interval: (*i*) for the same VM #2, $\sum_{t=t_1}^{t_3} \Phi_t \neq \Phi_T$, which violates *Additivity*; (*ii*) it produces different energy allocation for VM #2 and #3 over time $[t_1, t_2, t_3]$ while they satisfy the *Symmetry* in time T. From the example, we can see the Policy 2 that allocates the non-IT energy in proportion to VMs' IT energy is not self-consistent.

Policy 3 violates *Efficiency* and *Symmetry*. Considering two VMs (denoted as #1 and #2) , we have $\Phi_{1,j} = F_j(P_1+P_2) - F_j(P_2)$ and $\Phi_{2,j} = F_j(P_1+P_2) - F_j(P_1)$ according to Policy 3. Then the sum of shares is $\Phi_{1,j} + \Phi_{2,j} = 2F_j(P_1+P_2) - F_j(P_1) - F_j(P_2)$. It requires $F_j(P_1+P_2) = F_j(P_1) + F_j(P_2)$ to hold *Efficiency*. However, $F_j(\cdot)$ is non-linear as shown in Sec. II and it may have a static term. Hence the above equation cannot establish. Another explanation for Policy 3 is VM #1 and VM #2 join the non-IT unit j sequentially, and $\Phi_{1,j} = F_j(P_1) - F_j(0)$, $\Phi_{2,j} = F_j(P_1+P_2) - F_j(P_1)$. This violates *Symmetry*. For example, if $P_1 = P_2$, then VM #1 and VM #2 are exactly the same. But $\Phi_{1,j} \neq \Phi_{2,j}$ due to the non-linear function $F_j(\cdot)$. Actually, we can hardly distinguish which VM joins first when thousands of VMs co-exist in a datacenter, keeping performing start-up and shut-down operations. The second explanation for Policy 3 is not feasible in practice. So in the rest of the paper, Policy 3 refers to the first explanation. More importantly, the static energy is omitted during energy accounting according Policy 3, violating *Efficiency*.

Table III summarizes how they violate the four axioms. In Sec. VII, we will evaluate the above policies quantitatively.

While Shapley value satisfies all the axioms, applying it to our problem has two major challenges.

Challenge 1: We see from (3) that energy accounting based on Shapley value requires the value of $F_j(\sum_{k\in\mathcal{X}} P_k)$, for all $\mathcal{X} \subseteq \mathcal{N}_j$, i.e., the non-IT unit j's energy consumption when a subset of VMs are connected. However, we can only measure non-IT unit j's total energy consumption $P_j = F_j(\sum_{i\in\mathcal{N}_j} P_i)$.

Challenge 2: Shapley value requires an exponential number of calculations ($O(2^N)$) to get the non-IT energy share. This is because Shapley value averages over all the possible combinations of VMs in the system, resulting in an intolerable computational complexity (e.g., over 24 hours for only 20 VMs, whereas a real virtualized datacenter may have tens of thousands of VMs). It is impossible to directly apply Shapley value in datacenter's non-IT energy accounting, especially in real-time energy accounting scenarios (e.g., energy accounting per second).

V. LEAPS

A. *Quadratic Approximation*

To tackle the above challenges, we propose a novel Lightweight Energy Accounting Policy based on Shapley value, called LEAPS. More concretely, LEAPS leverages a quadratic function to approximate energy usage of each non-IT unit as follows:

$$F_j(x) = \begin{cases} 0, & \text{when } x \le 0 \\ a_j \cdot x^2 + b_j \cdot x + c_j, & \text{otherwise} \end{cases} \tag{4}$$

where x is the total IT energy of VMs served by non-IT unit j, and a_j, b_j, and c_j are modeling parameters that we learn and calibrate online as we measure the non-IT unit j's energy. Note that the quadratic approximation comes from real-world measurements in Sec. II and previous studies [15], [13]. Obviously, certain type of non-IT units (e.g., outside air cooling system) follow a cubic energy function. But as our deviation analysis will show later in Sec. V-B, the results of LEAPS, which leverages a quadratic function to approximate cubic function, is still fairly accurate compared to the original Shapley value approach. Note that the linear function can be treated as a special quadratic function whose $a_j = 0$, so we do not discuss it separately in the rest of paper.

When VM i has a zero IT energy (e.g., when it is shut down or put into sleep mode), its non-IT energy is clearly also zero, according to the null player axiom. So in the following theoretical derivation, we consider the case when VM i has a nonzero IT energy during an accounting period (e.g., every second).

When $\mathcal{X} = \emptyset$, by applying (4) into (3) and letting $|\mathcal{N}_j'| = n_j$, where $\mathcal{N}_j' \subseteq \mathcal{N}_j$ represents the set of VMs that have non-zero IT energy, we have

$$\begin{aligned} &\sum_{\mathcal{X}=\emptyset} \frac{|\mathcal{X}|!(|\mathcal{N}_j'| - |\mathcal{X}| - 1)!}{|\mathcal{N}_j'|!} \cdot [F_j(P_\mathcal{X} + P_i) - F_j(P_\mathcal{X})] \\ &= \frac{1}{n_j} \cdot (a_j P_i^2 + b_j P_i + c_j) \end{aligned} \tag{5}$$

When $\mathcal{X} \neq \emptyset$, by applying the quadratic function $F_j(x)$ into (3) and letting $|\mathcal{X}| = r_\mathcal{X}$ we have

$$\begin{aligned} &\sum_{\mathcal{X}\neq\emptyset, \mathcal{X}\subseteq\mathcal{N}_j'\setminus\{i\}} \frac{|\mathcal{X}|!(|\mathcal{N}_j'| - |\mathcal{X}| - 1)!}{|\mathcal{N}_j'|!} \\ &\qquad \cdot [F_j(P_\mathcal{X} + P_i) - F_j(P_\mathcal{X})] \\ &= \frac{2a_j P_i}{n_j!} \cdot \sum_{\mathcal{X}\neq\emptyset, \mathcal{X}\subseteq\mathcal{N}_j'\setminus\{i\}} [r_\mathcal{X}!(n_j - r_\mathcal{X} - 1)! P_\mathcal{X}] \\ &+ (a_j P_i^2 + bP_i) \cdot \sum_{\mathcal{X}\neq\emptyset, \mathcal{X}\subseteq\mathcal{N}_j'\setminus\{i\}} \frac{r_\mathcal{X}!(n_j - r_\mathcal{X} - 1)!}{n_j!}, \end{aligned} \tag{6}$$

where $P_\mathcal{X} = \sum_{k\in\mathcal{X}} P_k$. Note that over all nonempty subsets $\mathcal{X} \subseteq \mathcal{N}_j' \setminus \{i\}$ that have the same size/cardinality of u, there are $\binom{n_j-2}{u-1}$ subsets of $\mathcal{X} \setminus \{k\}$. Thus, in all the subsets $\mathcal{X} \subseteq \mathcal{N}_j' \setminus \{i\}$ whose size/cardinality equals to u, each VM k's IT energy P_k appears $\binom{n_j-2}{u-1} = \frac{(n_j-2)!}{(u-1)!(n_j-u-1)}$ times. Using this, we have

$$\begin{aligned} &\sum_{\mathcal{X}\neq\emptyset, \mathcal{X}\subseteq\mathcal{N}_j'\setminus i} [r_\mathcal{X}!(n_j - r_\mathcal{X} - 1)! P_\mathcal{X}] \\ &= \sum_{u=1}^{n_j-1} \sum_{\mathcal{X}, s.t., |\mathcal{X}|=u} [u!(n_j - u - 1)! P_\mathcal{X}] \\ &= \sum_{u=1}^{n_j-1} u(n_j-2)! \sum_{k\in\mathcal{N}_j'\setminus\{i\}} P_k \\ &= \frac{n_j!}{2} \sum_{k\in\mathcal{N}_j'\setminus\{i\}} P_k \end{aligned} \tag{7}$$

Similarly, there are $\binom{n_j-1}{u} = \frac{(n_j-1)!}{u!(n_j-u-1)!}$ non-empty subsets $\mathcal{X} \subseteq \mathcal{N}_j' \setminus \{i\}$ that have the same size/cardinality of u. Using this, we have

$$\begin{aligned} &\sum_{\mathcal{X}\neq\emptyset, \mathcal{X}\subseteq\mathcal{N}_j'\setminus\{i\}} \frac{r_\mathcal{X}!(n_j - r_\mathcal{X} - 1)!}{n_j!} \\ &= \sum_{u=1}^{n_j-1} \sum_{\mathcal{X}, s.t., |\mathcal{X}|=u} \frac{u!(n_j - u - 1)!}{n_j!} \\ &= \sum_{u=1}^{n_j-1} \frac{1}{n_j} = \frac{n_j - 1}{n_j} \end{aligned} \tag{8}$$

Next, by plugging (7) and (8) into (6), then plugging (5) and (6) into (3), the share of non-IT unit j's energy that LEAPS attributes to VM i is derived as

$$\Phi_{ij} = \begin{cases} 0, & \text{if } P_i = 0 \\ P_i \cdot [a_j \sum_{k\in\mathcal{N}_j'} P_k + b_j] + \frac{c_j}{n_j}, & \text{otherwise} \end{cases} \tag{9}$$

Firstly, if the non-IT power characteristic is indeed subject to a quadratic function, then LEAPS exactly equals to Shapley value. Second, LEAPS offers a closed-form expression of energy accounting with an interesting insight: the static energy of a non-IT unit is equally split among all the served VMs with non-zero energy, while the dynamic energy is attributed in proportion to VM's IT energy usage (since the term "$a_j \sum_{k\in\mathcal{N}_j'} P_k + b_j$" is the same for all VMs served by non-IT unit j). Compared with Shapley value, LEAPS is very easy to implement with negligible computation time, by combining two existing non-IT energy accounting policies (i.e., proportional for dynamic energy and equal for static energy).

It is important to note that LEAPS is derived from Shapley value and follows the allocation rule of Shapley value. The difference is that LEAPS leverages a quadratic function to approximate $F_j(\cdot)$ as the inputs of Shapley value.

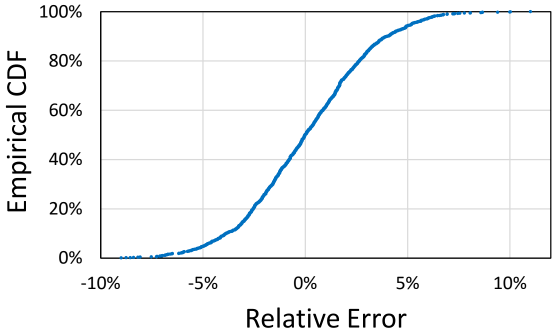


Fig. 4. $\delta_{P_{\mathcal{X}}}$ (normalized into relative error) distribution of UPS power loss.

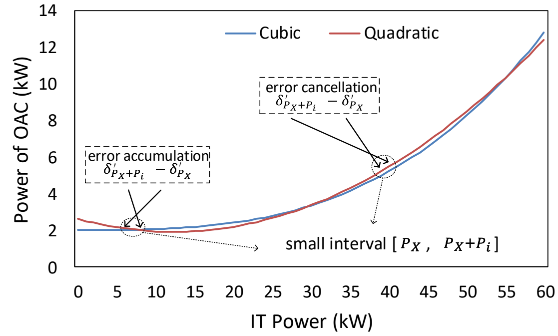


Fig. 5. An example of quadratic approximation ($F = 0.0046x^2 - 0.1126x + 2.5957$) to a cubic function ($F = 0.00005x^3 + 2$) from 0 to 60 kW.

Remark 1: In this paper, we use the least square fitting method to obtain a fitted quadratic function for each non-IT unit, even it has cubic power characteristic.

B. *Deviation of LEAPS from Shapley Value*

Though LEAPS follows the rule of Shapley value, it results in deviation from Shapley value due to the inputs. On the other hand, the Shapley value requires over 24 hours to calculate the results even for only 20 VMs. We can only compare LEAPS and Shapley value when VM scale is small. How do we quantify the deviation of LEAPS when VM scale is large? Next, we introduce a solution to the above challenge.

Suppose the difference between an approximated quadratic function $F_j(x)$ and the real power of non-IT unit $F_j^{'}(x)$ is δ_x when IT power is x, then $F_j^{'}(x)$ can be denoted as

$$F_j^{'}(x) = a_j \cdot x^2 + b_j \cdot x + c_j + \delta_x. \tag{10}$$

By plugging $F_j^{'}(x)$ into Shapley value and following the derivations in Sec. V, the original Shapley value $\Phi_{ij}^{'}$ can be easily derived as follows:

$$\Phi_{ij}^{'} = P_i \cdot [a_j \textstyle\sum_{k\in\mathcal{N}_j^{'}} P_k + b_j] + \frac{c_j}{|\mathcal{N}_j^{'}|} + \displaystyle\sum_{\mathcal{X}\subseteq\mathcal{N}_j^{'}\setminus\{i\}} \frac{|\mathcal{X}|!(|\mathcal{N}_j^{'}|-|\mathcal{X}|-1)!}{|\mathcal{N}_j^{'}|!} \cdot (\delta_{P_{\mathcal{X}}+P_i} - \delta_{P_{\mathcal{X}}}). \tag{11}$$

Obviously, the difference (denoted as Δ) between LEAPS and original Shapley value can be denoted as

$$\Delta = \sum_{\mathcal{X}\subseteq\mathcal{N}_j^{'}\setminus\{i\}} \frac{|\mathcal{X}|!(|\mathcal{N}_j^{'}|-|\mathcal{X}|-1)!}{|\mathcal{N}_j^{'}|!} \cdot (\delta_{P_{\mathcal{X}}+P_i} - \delta_{P_{\mathcal{X}}}). \tag{12}$$

Δ still has a computation complexity of $O(2^n)$. But from Equation (5) and (8), we can easily know

$$\sum_{\mathcal{X}\subseteq\mathcal{N}_j^{'}\setminus\{i\}} \frac{|\mathcal{X}|!(|\mathcal{N}_j^{'}|-|\mathcal{X}|-1)!}{|\mathcal{N}_j^{'}|!} = 1. \tag{13}$$

Δ can be interpreted as follows: $0 < |\mathcal{X}|!(|\mathcal{N}_j^{'}| - |\mathcal{X}| - 1)!/|\mathcal{N}_j^{'}|! < 1$ is a weight, Δ a weighted average of all possible $(\delta_{P_{\mathcal{X}}+P_i} - \delta_{P_{\mathcal{X}}})$. Note that the sum of wight equals to 1 according to Equation (13).

Hence, evaluating Δ can be transformed into a sampling and statistical problem: each $P_{\mathcal{X}}$ is a sampling location, ($\delta_{P_{\mathcal{X}}}$, $\delta_{P_{\mathcal{X}}+P_i}$) is a pair of sampling values, and the question is: when the sampling size is $2^{|\mathcal{N}_j^{'}|-1}$, how large is the weighted average of all sampled $(\delta_{P_{\mathcal{X}}+P_i} - \delta_{P_{\mathcal{X}}})$? Considering $\delta_{P_{\mathcal{X}}+P_i}$ and $\delta_{P_{\mathcal{X}}}$ may be negative or positive, $(\delta_{P_{\mathcal{X}}+P_i} - \delta_{P_{\mathcal{X}}})$ can be an accumulation or cancellation. So the value of $\delta_{P_{\mathcal{X}}}$ and $\delta_{P_{\mathcal{X}}+P_i}$ determines Δ. Next we discuss two type errors (**uncertain and certain error**) that constitute $\delta_{P_{\mathcal{X}}}$ and $\delta_{P_{\mathcal{X}}+P_i}$.

For the non-IT units that have quadratic power characteristic, $\delta_{P_{\mathcal{X}}}$ and $\delta_{P_{\mathcal{X}}+P_i}$ are caused by their power instability. From Fig. 2, we can see not all of the measured results of UPS perfectly lies on the approximated quadratic curve. Such $\delta_{P_{\mathcal{X}}}$ and $\delta_{P_{\mathcal{X}}+P_i}$ are usually uncertain, but small. After analyzing the measurement results of UPS in Fig. 2, we find $\delta_{P_{\mathcal{X}}}$ and $\delta_{P_{\mathcal{X}}+P_i}$ (normalized into relative error), are approximately subject to a normal distribution as shown in Fig. 4 , where $\mu = 0, \sigma = 0.023$ (i.e., around 95% of the relative errors $< 4.6\%$). For ease of reference, we call such $\delta_{P_{\mathcal{X}}}$ and $\delta_{P_{\mathcal{X}}+P_i}$ **uncertain error**. The uncertain errors are naturally small. Statistically speaking, Δ would be still small after all $(\delta_{P_{\mathcal{X}}+P_i} - \delta_{P_{\mathcal{X}}})$ take a weighted average.

Besides uncertain error, non-IT units with cubic power characteristic have certain error. Fig. 5 shows an example of quadratic approximation to a cubic function. **Certain error** means the difference between fitted quadratic function and the cubic one (denoted as $\delta_{P_{\mathcal{X}}}^{'}$, $\delta_{P_{\mathcal{X}}+P_i}^{'}$ in Fig. 5). Note that (*i*) $[P_{\mathcal{X}}, P_{\mathcal{X}} + P_i]$ is a small interval as one VM's power P_i is relatively small (about 0 to 300 W) compared with the total IT power (60 kW) on the x-axis. (*ii*) From Fig. 5, we can see the certain error accumulation only occurs when $[P_{\mathcal{X}}, P_{\mathcal{X}} + P_i]$ contains intersection points of cubic curve and fitted quadratic one. Due to (*i*), if we randomly pick a sampling location $P_{\mathcal{X}}$ on the x-axis, the probability that $(\delta_{P_{\mathcal{X}}+P_i}^{'} - \delta_{P_{\mathcal{X}}}^{'})$ is a cancellation is much higher than accumulation. (*iii*) In addition, the certain errors near intersection points are small, leading to a small error accumulation, and due to (*i*), $(\delta_{P_{\mathcal{X}}+P_i}^{'} - \delta_{P_{\mathcal{X}}}^{'})$ is very small if it is a cancellation. Statistically, the weighted average of $(\delta_{P_{\mathcal{X}}+P_i}^{'} - \delta_{P_{\mathcal{X}}}^{'})$ would be small.

Overall, Δ is small based on the above analysis. For further verification, we next evaluate Δ from the statistical point of view quantitatively. Note that the uncertain errors can be simulated using the normal distribution and certain error can be calculated. We can also calculate original Shapley value with $F_j^{'}(x)$ by adding the certain and uncertain errors, regardless of the high computation complexity.

VI. IMPLEMENTATION

A. Modeling VM Power

Note that VM power modeling is not the focus of this paper, and it has been widely studied in previous works [5], [27], [28], [29], [30], [31]. Here, we only introduce a solution to provide VM power trace for non-IT energy evaluation. The basic idea of VM energy modeling is to establish a power model which links the power consumption of a VM with its resource metrics (e.g., CPU utilization). Remarkably, the most common power model is the linear one, which is lightweight with over 90% of accuracy [4], [5]. In short, the equation form of a VM i's linear power model is as follows:

$$P_i = C_{cpu,i} \times u_{cpu,i} + C_{mem,i} \times u_{mem,i} + C_{disk,i} \times u_{disk,i} + C_{nic,i} \times u_{nic,i} \quad (14)$$

P_i refers to the predicted power consumption of VM i. $C_{cpu,i}$, $C_{mem,i}$, $C_{disk,i}$ and $C_{nic,i}$ are the coefficient of CPU, memory, disk and network interface card, respectively and $u_{cpu,i}$, $u_{mem,i}$, $u_{disk,i}$ and $u_{nic,i}$ are the utilization of the corresponding components.

Similarly, the linear power model is also effective for physical machine [4]. Considering the VMs in a datacenter may have different configuration, as well as different coefficients of power model, the power model training process of VMs will be tedious. Thus we develop a flexible power estimation method for VMs based on physical machine's resource utilization for brevity. Usually, the configuration of the physical machines is fixed, hence it only needs a one time model building phase to extract power consumption coefficient of their components. Then we re-scale the resource utilization of VMs by

$$u_{cpu,i}^{'} = \frac{u_{cpu,i} \times c}{C}, \quad u_{mem,i}^{'} = \frac{u_{mem,i} \times m}{M},$$
$$u_{disk,i}^{'} = \frac{u_{disk,i} \times d}{D}, \quad u_{nic,i}^{'} = \frac{u_{nic,i} \times n}{N}, \quad (15)$$

where c, m, d, n are the CPU cores, memory space, disk space and bandwidth allocated to VM i, and C, M, D, N are the maximum resource number of corresponding components in the physical machine. We train a physical machine's power model according to (14). Then we can obtain each VM's power consumption by applying the re-scaled utilization into physical machine's power model.

B. Parameter Setup

Table IV summarizes the values of parameters set in our evaluation. We setup the parameters based on the following considerations:

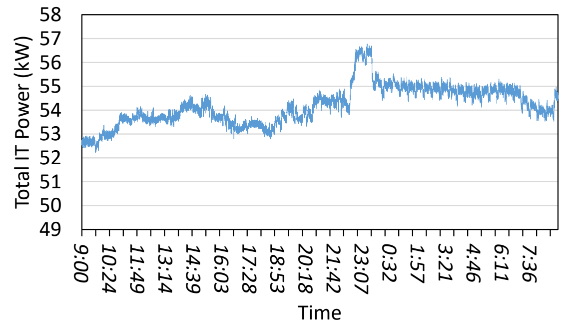
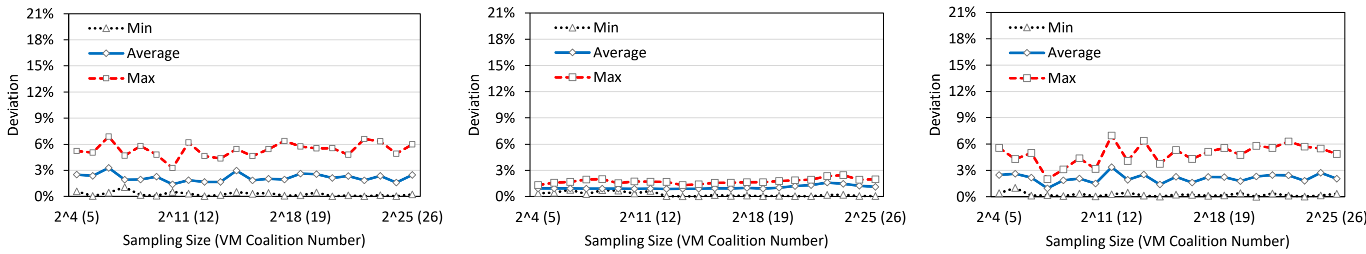


Fig. 6. IT power trace of the datacenter in a day (sampling interval: 1 second).

TABLE IV
PARAMETER SETTINGS OF OUR EXPERIMENTS

Accounting Interval		1 second
UPS Power setting		$F(x) = 0.0003x^2 + 0.0205x + 2.8628$
OAC Power Setting	Cubic	$F(x) = 0.0005x^3 + 2$, 30°C [14]
	Quadratic Fitting	$F(x) = 0.0457x^2 - 0.1255x$ $+5.9566$, $0 < x < 60$
Normal Distribution of Uncertain error	UPS	$\mu = 0, \sigma = 0.023$
	OAC	

- Energy accounting interval: We set the time interval between two non-IT energy accounting operations as 1 seconds. In previous works, it is called real-time power accounting [4].
- IT Power trace: We use Fluke 1378 three-phase power logger to record the IT power load in a day from our datacenter as shown in Fig. 6.
- We set 500 VMs running when sampling the IT power trace in Fig. 6.
- Non-IT units: We evaluate two representative non-IT units: the UPS and outside air cooling system, whose power consumption are quadratic and cubic, respectively.
- UPS power setting: We use the measurement results of the UPS in our datacenter, whose power function is listed in Sec. II.
- OAC power setting: As there is no OAC system in our datacenter, we use a cubic power function of OAC from previous study [14]: $F(x) = 0.0005x^3$, when outside temperature is 30 °C, and the approximated quadratic function is $F(x) = 0.0457x^2 - 0.1255x + 5.9566, 0 < x < 60$.
- Uncertain and certain error: As analyzed in Sec. V-B, uncertain errors are approximatively subject to a normal distribution, where $\mu = 0$ and $\sigma = 0.023$. According to it, we simulate the uncertain errors. For the OAC system, the certain error can be calculated using the above cubic and approximated quadratic function. Besides, OAC also has uncertain errors (the same as UPS). Accordingly, we can calculate the original Shapley value for UPS and OAC, by adding certain and uncertain errors.



(a) UPS (uncertain error) (b) OAC (only certain error) (c) OAC (certain + uncertain error)

Fig. 7. Deviation variation of LEAPS with different sampling size.

TABLE V
COMPUTATION TIME COMPARISON

VM Number	*Computation Time of LEAPS*	*Computation Time of Shapley value*
5	0.004 ms	0.016 ms
10	0.009 ms	33.59 s
15	0.014 ms	23 min 2 s
20	0.017 ms	$>$ 1 day
10000	2.1 ms	intolerable computation time

VII. EVALUATION

In this section, we demonstrate that LEAPS is reasonably accurate compared to the original Shapley value.

A. *Computation Time and Accuracy of LEAPS*

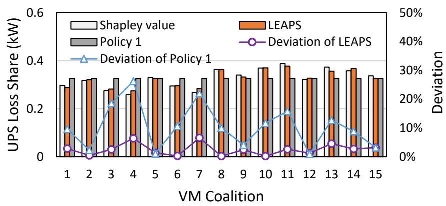
Computation Time: We run LEAPS and Shapley value on the same server with Intel Xeon E5-2650 V4 CPU. Table. V shows the computation time of LEAPS and Shapley value for different VM scales. We see that Shapley value becomes computationally prohibitive for real-time energy accounting when the VMs' number is 10, and it takes over 1 day for 20 VMs. Thus, it is infeasible to be implemented in a datacenter with tens of thousands of VMs. However, LEAPS only takes 2.1 ms to account even 10000 VMs' non-IT energy.

Accuracy: we can calculate the original Shapley value when VM scale is very small and compare LEAPS with it. But how can we evaluate LEAPS's accuracy when VM scale is large (e.g., 500 VMs)? From the analysis in Sec. V-B, we know $\delta_{P_{\mathcal{X}}}$ and $\delta_{P_{\mathcal{X}}+P_i}$ consist of certain and uncertain errors, and evaluating deviation of LEAPS is transformed into a sampling and statistical problem: each $P_{\mathcal{X}}$ is a sampling location, ($\delta_{P_{\mathcal{X}}}$, $\delta_{P_{\mathcal{X}}+P_i}$) is a pair of sampling values, and the question is: when sampling size is $2^{|\mathcal{N}_j^{'}|-1}$, how large is the weighted average of all sampled $(\delta_{P_{\mathcal{X}}+P_i} - \delta_{P_{\mathcal{X}}})$? Note that the sum of weight is 1, and each weight is a value between (0, 1).

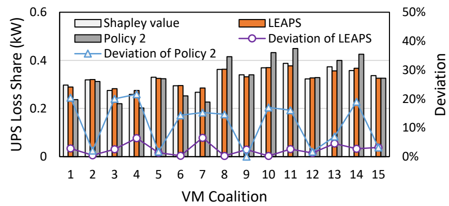
Accordingly, we first randomly divide the 500 VMs into 5 coalitions when total IT power is 54.872 kW, and calculate the non-IT energy accounting results from Shapley value and LEAPS for the 5 coalitions, respectively. Then we extend the number of divided coalitions from 5 to 26 step by step, and run a simulation for a month to account the non-IT energy for the different number of VM coalitions. Then we compare the results of LEAPS and Shapley value. Fig. 7 shows the deviation of LEAPS from Shapley value when VM coalition number varies from 5 to 26. Though the number of coalitions is small, the sampling size grows exponentially from 32 to over 33.5 million. In other words, we sample over 33.5 million pairs of ($\delta_{P_{\mathcal{X}}}$, $\delta_{P_{\mathcal{X}}+P_i}$) in the range of [0, 54.872] when VM coalition number is 26. As shown in Fig. 7(a) and Fig. 7(c), we can see that, when the sampling size varies in a range of [32, 64, 128, ..., 33554432], the weighted average of $(\delta_{P_{\mathcal{X}}+P_i} - \delta_{P_{\mathcal{X}}})$ (i.e., deviation of LEAPS) still maintains small within an average relative error less than 3.28% and 3.37%, and a maximum relative error of 6.88% and 6.97% for UPS and OAC, respectively. In addition, the effect of certain error on LEAPS' accuracy is small within an average error less than 1.59% as shown in Fig. 7(b). The above results are consistent with our analysis in Sec. V-B: for the uncertain errors which are naturally small, it is still small after they take a weighted average; for the certain errors, most of $(\delta_{P_{\mathcal{X}}+P_i} - \delta_{P_{\mathcal{X}}})$ are error cancellation, so their weighted average is still small. Base on the above results and analysis in Sec. V-B, it is reasonable to say the weighted average of $(\delta_{P_{\mathcal{X}}+P_i} - \delta_{P_{\mathcal{X}}})$ still remains small, even when VM scale is 500 with 2^{500} samples. Hence, from the statistical point of view, we demonstrate that compared with Shapley value, LEAPS is reasonably accuracy within a maximum relative error of 6.97%.

B. *LEAPS Against Other Policies*

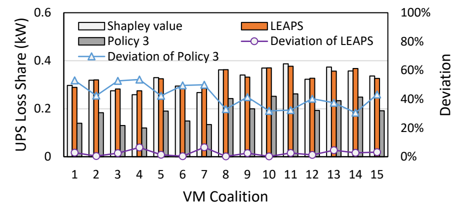
Due to the high computation complexity of Shapley value, we randomly divide the VMs into 15 coalitions when IT power is 54.872 kW, and account their non-IT energy using different policies (i.e., Policy 1∼3, LEAPS and Shapely value) for comparison. Fig. 8 and Fig. 9 show the results of different policies compared with Shapley value. We can see existing policies have large deviations from Shapley value. On the contrary, LEAPS only has a maximum error of 6.53% for UPS and 5.92% for OAC. Policy 1 equally splits the non-IT energy, which is obviously unfair. The biggest difference between LEAPS and Policy 2 is that LEAPS splits static energy among VMs with non-zero IT energy and OAC has no static energy consumption. Thus, Policy 2 has a similar result to LEAPS when accounting OAC energy. Policy 3 is based on marginal energy increment and the static energy is omitted during energy accounting. So it allocates much less UPS loss compared with other policies. For OAC that has no



(a) LEAPS VS. Policy 1

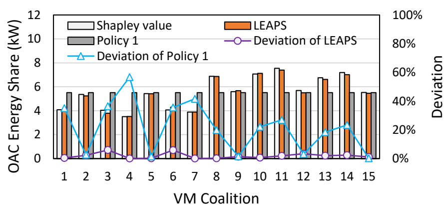


(b) LEAPS VS. Policy 2

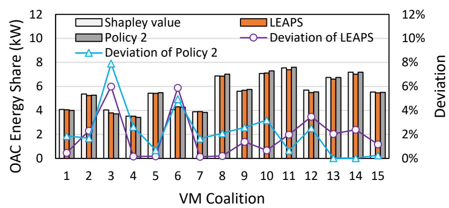


(c) LEAPS VS. Policy 3

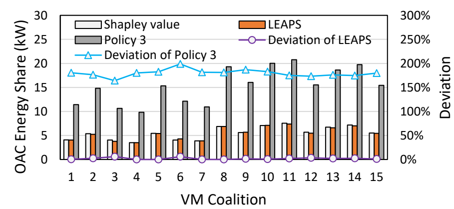
Fig. 8. UPS loss accounting result comparison of different policies.



(a) LEAPS VS. Policy 1



(b) LEAPS VS. Policy 2



(c) LEAPS VS. Policy 3

Fig. 9. OAC energy accounting result comparison of different policies.

static energy, it allocates much lager non-IT share to each VM due to the cubic energy growth of OAC.

VIII. Related Work

Non-IT Equipment Optimization: Efforts have been paid to improve the energy efficiency of non-IT equipments. [32] proposes a market-based approach to improve power infrastructure utilization. [33] studies a new power architecture to reduce energy cost in datacenter. The energy optimization of cooling system has also been widely investigated [34], [35], [36]. But the non-IT energy of individual tenant still remains ambiguous.

Datacenter Energy Accounting: Datacenter energy accounting has received much attention in recent years [37]. To reduce hardware cost, a power disaggregation approach is proposed in [4] to estimate the power consumption of individual servers in legacy datacenters. [5], [28] study model-based power metering for virtualized datacenters, with the goal of better utilizing the expensive power infrastructure. Motivated by nested virtualization, BITWATTS [31] proposes a process-aware model to account the power of nested VMs. [38] profiles power usage by different applications on hyper-threaded processors. To access more fine-grained power usage at component level, [7] constructs power proxies for CPU, at both chip level and core level, by utilizing specialized activity counters in the chip hardware. Further, [14] develops IT and non-IT energy models at system level to minimize datacenter energy usage, [15] proposes measurement-based power models for different non-IT units (e.g., UPS and cooling) in datacenters. In contrast, we focus on non-IT energy accounting in virtualized datacenters, which has not been well investigated.

Solutions Using Shapley: Shapley value has been used in other contexts such as energy accounting on mobile systems [26], virtual machine power accounting [27] and peak demand cost splitting across users in cloud datacenters [39]. Our work differs from [26], [27], [39] in that we study an orthogonal problem and propose a novel low-complexity method with little to zero implementation overhead. Our method exploits the unique characteristics of datacenter non-IT energy model, and also differs from the generic random sampling-based fast Shapley value calculation that may yield large errors [40]. Finally, note that fair multi-resource allocation in computer systems [41], [42] has a fundamentally different goal than our work: it aims at "fairly" improving system utilization by encouraging users' resource sharing. It emphasizes the "equality/balance" of users' resource allocation. Whereas we re-attribute non-IT energy to different VMs and fairness in our context, as supported by Shapley value theorem [17], [18], means satisfying the axioms in Sec. IV-B. It emphasizes "more pay for more work".

IX. Conclusion

In this paper, we study the power consumption characteristic of non-IT units in datacenter. Then we formulate the non-IT energy accounting problem and demonstrate Shapley value provides the theoretical ground truth. To overcome the high computation complexity of Shapley value, we propose LEAPS, which leverages a quadratic function to approximate the inputs of Shapley value. LEAPS is easy to implement and offers a new perspective on non-IT energy accounting: it allocates the dynamic energy in proportion to VM's IT energy and equally splits the static energy among running VMs. Further, our experiments show LEAPS is effective to approximate Shapley value within maximum relative error less than 6.97%. In addition, LEAPS may also be applied to those areas outside of non-IT energy, where the gain/cost grows quadratically, e.g., computational sprinting [43], [44].

REFERENCES

[1] G. Peace, http://www.greenpeace.org/international/en/press/releases/2017/Amazon-still-lags-behind-Apple-Google-in-Greenpeace-renewable-energy-report/.

[2] A. Inc., "Environmental responsibility report," *2017 Progress Report*.

[3] Akamai, https://www.akamai.com/uk/en/about/news/press/2016-press/akamai-advances-sustainability-initiative-for-renewable-energy.jsp, 2016.

[4] G. Tang, W. Jiang, Z. Xu, F. Liu, and K. Wu, "Zero-cost, fine-grained power monitoring of datacenters using non-intrusive power disaggregation," in *Proceedings of the 16th Annual Middleware Conference (Middleware'15)*. ACM, 2015.

[5] A. Kansal, F. Zhao, J. Liu, N. Kothari, and A. A. Bhattacharya, "Virtual machine power metering and provisioning," in *Proceedings of the ACM symposium on Cloud computing (SoCC'10)*. ACM, 2010.

[6] X. Feng, R. Ge, and K. W. Cameron, "Power and energy profiling of scientific applications on distributed systems," in *Proceedings of Parallel and Distributed Processing Symposium (IPDPS'05)*. IEEE, 2005.

[7] W. Huang, C. Lefurgy, W. Kuk, A. Buyuktosunoglu, M. Floyd, K. Rajamani, M. Allen-Ware, and B. Brock, "Accurate fine-grained processor power proxies," in *Proceedings of Annual IEEE/ACM International Symposium on Microarchitecture (MICRO'12)*. IEEE Computer Society, 2012.

[8] G. Grid, "Pue: A comprehensive examination of the metric," *White Paper*, 2012.

[9] U. Institute, "2014 Data Center Industry Survey," https://journal.uptimeinstitute.com/2014-data-center-industry-survey/, 2014.

[10] Asetek, https://www.asetek.com/data-center/benefits/data-center-demands/.

[11] CoolIT, "Rack dclc product guide," http://coolit2017.qt4egaquh7.maxcdn-edge.com/wp-content/uploads/2017/09/rack-dclc-product-guide.pdf.

[12] Y. L. Du Su, "Greenmap: mapreduce with ultra high efficiency power delivery," 2015.

[13] R. L. Sawyer, "Making large ups systems more efficient," *Schneider Electric White Papers 108, Revision 3*, 2012.

[14] Z. Liu, Y. Chen, C. Bash, A. Wierman, D. Gmach, Z. Wang, M. Marwah, and C. Hyser, "Renewable and cooling aware workload management for sustainable data centers," in *Proceedings of ACM SIGMETRICS Performance Evaluation Review (SIGMETRICS'12)*. ACM, 2012.

[15] S. Pelley, D. Meisner, T. F. Wenisch, and J. W. VanGilder, "Understanding and abstracting total data center power," in *Proceedings of Workshop on Energy-Efficient Design (WEED'09)*, 2009.

[16] Q. Zhang and W. Shi, "Ups-aware workload placement in enterprise data centers," *IEEE Computer Magazine*, 2015.

[17] L. S. Shapley, "A value for n-person games," *Contributions to the Theory of Games*, 1953.

[18] R. Gopalakrishnan, J. R. Marden, and A. Wierman, "Potential games are necessary to ensure pure nash equilibria in cost sharing games," *Mathematics of Operations Research*, 2014.

[19] M. A. Islam, S. Ren, and A. Wierman, "Exploiting a thermal side channel for power attacks in multi-tenant data centers," in *Proceedings of the SIGSAC Conference on Computer and Communications Security (CCS)*. ACM, 2017.

[20] F. Corporation, "Fluke 1736 and 1738 Three-Phase Power Loggers," http://media.fluke.com/documents/6006033a-173x-power-logger-ds-w.pdf, 2015.

[21] W. Zheng and X. Wang, "Data center sprinting: Enabling computational sprinting at the data center level," in *Proceedings of International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2015.

[22] C. Wang, B. Urgaonkar, A. Gupta, L. Y. Chen, R. Birke, and G. Kesidis, "Effective capacity modulation as an explicit control knob for public cloud profitability," in *Proceedings of International Conference on Autonomic Computing (ICAC)*. IEEE, 2016.

[23] M. A. Islam, H. Mahmud, S. Ren, and X. Wang, "Paying to save: Reducing cost of colocation data center via rewards," in *Proceedings of International Symposium on High Performance Computer Architecture (HPCA'15)*. IEEE, 2015.

[24] R. Stanojevic, N. Laoutaris, and P. Rodriguez, "On economic heavy hitters: shapley value analysis of 95th-percentile pricing," in *Proceedings of ACM SIGCOMM conference on Internet measurement (IMC'10)*. ACM, 2010.

[25] V. Misra, S. Ioannidis, A. Chaintreau, and L. Massoulié, "Incentivizing peer-assisted services: a fluid shapley value approach," in *Proceedings of ACM SIGMETRICS Performance Evaluation Review (SIGMETRICS'10)*. ACM, 2010.

[26] M. Dong, T. Lan, and L. Zhong, "Rethink energy accounting with cooperative game theory," in *Proceedings of the 20th annual international conference on Mobile computing and networking (MobiCom'14)*. ACM, 2014.

[27] W. Jiang, F. Liu, G. Tang, K. Wu, and H. Jin, "Virtual machine power accounting with shapley value," in *Proceedings of International Conference on Distributed Computing Systems (ICDCS),*. IEEE, 2017.

[28] H. Lim, A. Kansal, and J. Liu, "Power budgeting for virtualized data centers," in *Proceedings of USENIX Annual Technical Conference (USENIX ATC'11)*. USENIX, 2011.

[29] B. Krishnan, H. Amur, A. Gavrilovska, and K. Schwan, "Vm power metering: feasibility and challenges," *ACM SIGMETRICS Performance Evaluation Review*, 2011.

[30] A. E. H. Bohra and V. Chaudhary, "Vmeter: Power modelling for virtualized clouds," in *International Symposium on Parallel & Distributed Processing, Workshops and Phd Forum (IPDPSW)*, IEEE, 2010.

[31] M. Colmant, M. Kurpicz, P. Felber, L. Huertas, R. Rouvoy, and A. Sobe, "Process-level power estimation in vm-based systems," in *European Conference on Computer Systems (EuroSys)*, ACM, 2015.

[32] M. Islam, X. Ren, S. Ren, and A. Wierman, "A spot capacity market to increase power infrastructure utilization in multi-tenant data centers," in *Proceedings of International Symposium on High Performance Computer Architecture (HPCA)*. IEEE, 2018.

[33] L. Liu, C. Li, H. Sun, Y. Hu, J. Gu, T. Li, J. Xin, and N. Zheng, "Heb: Deploying and managing hybrid energy buffers for improving datacenter efficiency and economy," in *Proceedings of Annual International Symposium on Computer Architecture (ISCA)*. ACM, 2015.

[34] I. n. Goiri, T. D. Nguyen, and R. Bianchini, "Coolair: Temperature- and variation-aware management for free-cooled datacenters," in *Proceedings of International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*. ACM, 2015.

[35] I. Manousakis, I. n. Goiri, S. Sankar, T. D. Nguyen, and R. Bianchini, "Coolprovision: Underprovisioning datacenter cooling," in *Proceedings of ACM Symposium on Cloud Computing (SoCC)*. ACM, 2015.

[36] M. Skach, M. Arora, C.-H. Hsu, Q. Li, D. Tullsen, L. Tang, and J. Mars, "Thermal time shifting: Leveraging phase change materials to reduce cooling costs in warehouse-scale computers," in *Proceedings of Annual International Symposium on Computer Architecture (ISCA)*. ACM, 2015.

[37] A. Noureddine, R. Rouvoy, and L. Seinturier, "A review of energy measurement approaches," *ACM SIGOPS Operating Systems Review*, 2013.

[38] Y. Zhai, X. Zhang, S. Eranian, L. Tang, and J. Mars, "Happy: Hyperthread-aware power profiling dynamically." in *Proceedings of USENIX Annual Technical Conference (ATC'14)*. USENIX, 2014.

[39] N. Nasiriani, C. Wang, G. Kesidis, B. Urgaonkar, L. Y. Chen, and R. Birke, "On fair attribution of costs under peak-based pricing to cloud tenants," *ACM Transactions on Modeling and Performance Evaluation of Computing Systems (TOMPECS'16)*, 2016.

[40] J. Castro, D. Gómez, and J. Tejada, "Polynomial calculation of the shapley value based on sampling," *Computers & Operations Research*, 2009.

[41] A. Ghodsi, M. Zaharia, B. Hindman, A. Konwinski, S. Shenker, and I. Stoica, "Dominant resource fairness: Fair allocation of multiple resource types." in *Proceedings of USENIX Symposium on Networked Systems Design and Implementation (NSDI'11)*, 2011.

[42] T. Bonald and J. Roberts, "Multi-resource fairness: Objectives, algorithms and performance," in *Proceedings of ACM SIGMETRICS Performance Evaluation Review (SIGMETRICS'15)*. ACM, 2015.

[43] S. Ren, Y. He, S. Elnikety, and K. S. McKinley, "Exploiting processor heterogeneity in interactive services," in *Proceedings of the International Conference on Autonomic Computing (ICAC)*. USENIX, 2013.

[44] N. Morris, S. M. Renganathan, C. Stewart, R. Birke, and L. Chen, "Sprint ability: How well does your software exploit bursts in processing capacity?" in *Proceedings of International Conference on Autonomic Computing (ICAC)*. IEEE, 2016.