

Energy-aware Dynamic Pricing Model for Cloud Environments

Peini Liu¹, Gusseppe Bravo¹, and Jordi Guitart^{1,2}

¹ Barcelona Supercomputing Center (BSC), Barcelona, Spain

² Universitat Politècnica de Catalunya (UPC), Barcelona, Spain
{peini.liu,gussepe.bravo,jordi.guitart}@bsc.es

Abstract. Energy consumption is a critical operational cost for Cloud providers. However, as commercial providers typically use fixed pricing schemes that are oblivious about the energy costs of running virtual machines, clients are not charged according to their actual energy impact. Some works have proposed energy-aware cost models that are able to capture each client's real energy usage. However, those models cannot be naturally used for pricing Cloud services, as the energy cost is calculated after the termination of the service, and it depends on decisions taken by the provider, such as the actual placement of the client's virtual machines. For those reasons, a client cannot estimate in advance how much it will pay. This paper presents a pricing model for virtualized Cloud providers that dynamically derives the energy costs per allocation unit and per work unit for each time period. They account for the energy costs of the provider's static and dynamic energy consumption by sharing out them according to the virtual resource allocation and the real resource usage of running virtual machines for the corresponding time period. Newly arrived clients during that period can use these costs as a baseline to calculate their expenses in advance as a function of the number of requested allocation and work units. Our results show that providers can get comparable revenue to traditional pricing schemes, while offering to the clients more proportional prices than fixed-price models.

Keywords: Pricing Model · Energy Consumption · Cloud Computing

1 Introduction

Cloud computing has consolidated as a paradigm for the on-demand provisioning of computing resources to end users over the Internet. These services are executed in virtual machines (VMs) hosted in large-scale data centers, which have become greedy consumers of energy to provide those services. Greenpeace [1] estimates that data centers energy use can grow up to 1012 billion kWh by 2020, which is a 3x increment regarding their energy consumption in 2007. The cost of this enormous amount of energy has turned into the primary cost driver for data centers. Belady [6] estimates that the annual amortized energy costs in a data center for a single server exceeded the cost of the server itself in 2008. Consequently, any cost model for the Cloud should be energy-aware.

Cloud computing was originally devised as a utility computing paradigm, where the VMs had to be offered to the end users in a pay-as-you-use manner, i.e. the user pays only for the resources really consumed [8], such as any other utility service like water and electricity. However, commercial Cloud providers [4, 12, 14] typically charge their clients in a pay-as-you-go manner, i.e. the user pays a fixed value per unit of time for the VMs, whether he is using them or not [8]. Whereas those fixed prices encompass the operational costs of the provider, they are oblivious about the real energy cost to run each specific VM.

We claim that Cloud providers must offer an energy-aware and proportional dynamic pricing model to their users. Prices must be calculated dynamically because both the energy consumed in the data center (which depends on the number of clients and the amount of resources each of them uses) and the price of that energy vary over time (e.g. in Spain the energy price varies per hour [15]). Prices must be proportional so that users are charged according to their actual energy impact, i.e. clients using more energy should pay more.

Some works [2, 3, 8, 11, 13] have proposed cost models for VMs that can account for their individual energy impact. However, this energy cost is calculated after the termination of the VM, once the full energy usage profile of the VM and the provider's placement decisions about that VM are known. This impedes some of the advantages of fixed prices models discussed before, such as predictability, i.e. clients can know the cost of running their VMs before running them because it only depends on the client's behavior, and fairness, i.e. two identical VMs launched at the same time and with the same duration will pay the same.

This paper presents a pricing model for virtualized Cloud providers that is energy-aware, proportional, predictable, and fair. Our model builds upon the concepts of Allocation Units, which quantify the amount of virtual resources that are allocated to the VMs, and Work Units, which quantify the amount of work executed by using those resources. Our model dynamically derives the energy costs per Allocation Unit and per Work Unit for each time period. They account for the costs of the providers static and dynamic energy consumption by sharing out them according to the virtual resource allocation and the real resource usage of running VMs for the corresponding time period. Newly arrived clients during that period can use these costs as a baseline to calculate their expenses in advance as a function of the number of requested allocation and work units.

2 Related Work

Pricing models in Cloud Computing have been broadly classified as subscription-based (clients reserve resources in advance for a specific period of time by paying a fixed price up-front), pay-per-use (resources are provided on-demand and clients are charged a fixed price per unit of time on usage basis), and hybrid (combination of subscription-based and pay-per-use) [10]. However, commercial Cloud providers might classify their prices differently depending on their customers' requirements. For example, Amazon [4] offers On-Demand, Reserved, and Spot Pricing Instances, being the former the most popular among the clients. Other

providers such as Azure and Rackspace support similar pricing schemes [12, 14]. None of these commercial providers consider the real energy cost to run the VMs when charging the clients.

Some works [2, 3, 8, 11, 13] have proposed cost models for VMs that account for their individual energy impact, but this is calculated after the termination of the VM, thus a client cannot estimate in advance how much it will pay.

Aldossary and Djemame [2] proposed a pricing model charges the customer based on the actual resource usage per unit including the energy consumption. Their model distributes the dynamic energy among the VMs according to their utilization, but the static energy is distributed evenly among VMs, independently of their size. Furthermore, they use the average power to calculate the energy consumption, which is not very accurate when resource usage fluctuates.

Hinz et al. [8] presented a cost model which accounts for the individualized energy cost for each VM according to its CPU and network usage. As a novelty, it includes also a shared cost from common hypervisor management operations, which is proportionally distributed among VMs according to their number of virtual processors (as they do also with the static energy).

Kurpicz et al. [11] presented a model for energy-proportional accounting for VMs which determines their dynamic energy costs by using the real utilization of the resources and divides the static energy costs proportionally to the number of virtual processors of the VM. To offer some cost predictability to users, the model reports a lower and an upper bound of the VM total cost, but these bounds are very coarse-grained.

3 Problem Statement

Our purpose is to define a pricing scheme that determines how much a given virtual machine j will pay if it runs for a time period of D hours. We define this cost as $C_{VM}^j(D)$. As listed in the top part of Table 1, the user must only provide as inputs the number of requested Allocation Units by the VM and the number of Work Units to be executed by the VM.

We define the number of Allocation Units of a VM j (AU_{VM}^j) as the product of its number of virtual processors ($VCPU_{VM}^j$) and its amount of memory (RAM_{VM}^j), which are normalized with respect to the capabilities of an Amazon m1.small VM (i.e. $VCPU_{VM}^{M1} = 1$ core and $RAM_{VM}^{M1} = 1.7\text{GB}$). According to this, AU_{VM}^j is calculated as shown in Eq. (1). The number of virtual processors and the amount of memory of VMs are the critical parameters that determine the number of VMs that can be allocated in a physical host.

$$AU_{VM}^j = \frac{VCPU_{VM}^j}{VCPU_{VM}^{M1}} \cdot \frac{RAM_{VM}^j}{RAM_{VM}^{M1}} \quad (1)$$

We define the number of Work Units of a VM j (WU_{VM}^j) as the number of millions of instructions to be executed by the VM.

Table 1: Parameters used by the model.

Symbol	Description
V_{VM}^j	Number of virtual processors of VM j
RAM_{VM}^j	GB of RAM of VM j
AU_{VM}^j	Number of Allocation Units needed by VM j (calculated as in Eq. (1))
WU_{VM}^j	Number of Work Units to be executed by VM j
$E_{DC}(t)$	DC total energy consumption during time period t (in Joules)
$ES_{DC}(t)$	DC static energy consumption during time period t (in Joules)
$ED_{DC}(t)$	DC dynamic energy consumption during time period t (in Joules)
$E_H^i(t)$	Energy consumption of host i during time period t (in Joules)
$EI_H^i(t)$	Energy consumption of host i during time period t when idle (in Joules)
$AU_{DC}(t)$	Number of awarded Allocation Units in the DC during time period t
$WU_{DC}(t)$	Number of Work Units executed in the DC during time period t
$AU_H^i(t)$	Number of awarded Allocation Units in host i during time period t
$WU_H^i(t)$	Number of Work Units executed in host i during time period t
$E_{AU}(t)$	Energy consumption for each AU during period t (in Joules)
$E_{WU}(t)$	Energy consumption for each WU during period t (in Joules)
$C_{AU}(t)$	Cost of energy consumed per AU during period t (in €/kWh)
$C_{WU}(t)$	Cost of energy consumed per WU during period t (in €/kWh)
PI_H^i	Average power consumption of host i when idle (in Watts)
$MIPS_H^i$	Performance of host i (in Millions of Instructions per second)
$P_H^i(t)$	Instantaneous power consumption of host i at time t (in Watts)
U_H^i	Instantaneous CPU utilization of host i at time t ($\in [0, 1]$)
$N(t)$	Number of active hosts during time period t
$VMs_H^i(t)$	Amount of VMs on host i during time period t
$Price(t)$	Energy price during time period t (in €/kWh)
T_S	Time elapsed between two samples (in seconds)
$NS(t)$	Number of samples during time period t ; $NS(t) = t/T_S$

The provider offers the energy costs per Allocation and Work Unit for the time when the VM has been submitted (t_0). Next section describes how the provider accounts for these costs. The client can use them as a baseline to calculate their expenses in advance as a function of the number of requested Allocation and Work units, as follows: $C_{VM}^j(D) = (AU_{VM}^j \cdot C_{AU}(t_0) + WU_{VM}^j \cdot C_{WU}(t_0)) \cdot D$.

4 Pricing Model

To derive the energy costs per Allocation Unit and per Work Unit for each time period, our model calculates a number of parameters for each time period. They are introduced in the second part of Table 1. The model requires a number of input parameters that must be introduced by the provider. Some of these parameters are obtained by calibrating the data center in a profiling stage (see third part of Table 1), others are gathered periodically by monitoring the data center status (see fourth part of Table 1), and the rest are configuration parameters or other external inputs (see bottom of Table 1).

4.1 Energy consumption model

Our model must account first for the total energy consumption of the data center during each time period t ($E_{DC}(t)$), which can be calculated as the sum of the energy consumption from all the active hosts during that period, as shown in Eq. (2). The energy consumption of host i during time period t results from integrating all the power consumption of host i during that time period. Given that we do not have the continuous function describing that power consumption but a set of samples of its value, we do the calculation as presented in Eq. (3), where $P_H^i(t_k)$ is the k -th sample of the power consumption of host i .

$$E_{DC}(t) = \sum_{i=1}^{N(t)} E_H^i(t) \quad (2) \quad E_H^i(t) = \int_1^t P_H^i(t) dt = T_S \cdot \sum_{k=1}^{NS(t)} P_H^i(t_k) \quad (3)$$

The energy consumption of the data center ($E_{DC}(t)$) comprises both the static energy consumption ($ES_{DC}(t)$) due to keeping the hosts on and the dynamic energy consumption ($ED_{DC}(t)$) spent by all the running VMs to do their work, as shown next: $E_{DC}(t) = ES_{DC}(t) + ED_{DC}(t)$.

The static energy consumption of the data center ($ES_{DC}(t)$) is the sum of the idle energy consumption of all the active hosts. These come from their average idle power consumption during time period t as shown in Eq. (4).

$$ES_{DC}(t) = \sum_{i=1}^{N(t)} EI_H^i(t) = \sum_{i=1}^{N(t)} PI_H^i * t \quad (4)$$

The dynamic energy consumption of the data center ($ED_{DC}(t)$) is calculated from the total and the static energy consumption for the data center during time period t as follows: $ED_{DC}(t) = E_{DC}(t) - ES_{DC}(t)$.

4.2 Energy-aware cost model

The provider dynamically and periodically calculates its prices for an Allocation Unit (AU) and a Work Unit (WU) to reflect the variability in the electricity price and the energy consumption (due to changing workloads).

The static and dynamic energy costs are shared out according to the virtual resource allocation and the real resource usage of running VMs for the corresponding time period. This will result in a cost per Allocation Unit ($C_{AU}(t)$) and a cost per Work Unit ($C_{WU}(t)$) for that time period.

$C_{AU}(t)$ is calculated from the energy consumption per Allocation Unit during time period t as follows: $C_{AU}(t) = Price(t) \cdot E_{AU}(t)/3600000$. As shown in Eq. (5), $E_{AU}(t)$ is derived from the static energy consumption of the data center ($ES_{DC}(t)$), which was calculated in the previous section, and the number of awarded Allocation Units to all the running VMs in the data center ($AU_{DC}(t)$), which is the sum of all the Allocation Units awarded in all the active hosts during time period t ($AU_H^i(t)$). As shown in Eq. (6), it can be calculated as

the awarded Allocation Units from all the running VMs on each host i , where $VMs_H^i(t)$ represents the number of running VMs in host i during the time t and AU_{VM}^j is the number of Allocation Units of VM j .

$$E_{AU}(t) = \frac{ES_{DC}(t)}{AU_{DC}(t)} \quad (5) \quad AU_{DC}(t) = \sum_{i=1}^{N(t)} AU_H^i(t) = \sum_{i=1}^{N(t)} \sum_{j=1}^{VMs_H^i(t)} AU_{VM}^j \quad (6)$$

$C_{WU}(t)$ is calculated from the energy consumption per Work Unit during period t as follows: $C_{WU}(t) = Price(t) \cdot E_{WU}(t)/3600000$. As shown in Eq. (7), $E_{WU}(t)$ is derived from the dynamic energy consumption of the data center ($ED_{DC}(t)$), which was calculated in the previous section, and the number of Work Units executed in the data center ($WU_{DC}(t)$), which is the sum of the units executed in all the hosts during period t ($WU_H^i(t)$), as shown in Eq. (8).

$$E_{WU}(t) = \frac{ED_{DC}(t)}{WU_{DC}(t)} \quad (7) \quad WU_{DC}(t) = \sum_{i=1}^{N(t)} WU_H^i(t) \quad (8)$$

The number of Work Units that a host can execute depends on its performance capability. In this paper, we have defined a Work Unit as 1 million of instructions to be executed, and hence, we measure the performance of hosts using MIPS. According to this, the number of Work Units executed in host i during time period t ($WU_H^i(t)$) depends on the maximum performance of host i ($MIPS_H^i$) and the host utilization during time period t while running VMs (U_H^i). Without loss of generality, we measure the host utilization as its CPU utilization, since the CPU is the highest contributor to the power consumption of a host. However, our model could be easily extended to consider also the utilization of other resources. Given that we do not have the continuous function describing the CPU utilization of host i during time period t but a set of samples of its value, we calculate $WU_H^i(t)$ as shown in Eq. (9), where $U_H^i(t_k)$ is the k -th sample of the CPU utilization of host i (between 0 and 1).

$$WU_H^i(t) = MIPS_H^i \cdot \int_1^t U_H^i(t) dt = MIPS_H^i \cdot T_S \cdot \sum_{k=1}^{NS(t)} U_H^i(t_k) \quad (9)$$

5 Experiments and Evaluation

5.1 Experimental setup and workload

A data center comprising 200 high-performance hosts has been simulated with CloudSim-plus [7]. Each host consists of two Intel Xeon 8180M processors with 28 cores each, two 128GB PC42400U RAM DIMM, and two disks 2.5-3840GB-SATA, providing a computing performance of 143360 MIPS. Its idle and maximum rated power have been reported as 109.11 and 578.85 Watts, respectively [9]. The electricity fee paid by the data center is calculated according to the electricity price, which varies every hour [15]. The sampling period T_S is 5 minutes.

Table 2: Configuration of tasks

Task type	Amount	Instructions
1	500	15000000
2	500	20000000
3	500	28000000
4	500	35000000

Table 3: Configuration of VMs

VM type	vCPU	MIPS	RAM
m4.Large	2	5120	8 GB
m4.xLarge	4	10240	16 GB
m4.2xLarge	8	20480	32 GB
m4.4xLarge	16	40960	64 GB

The workload comprises 2000 tasks of 4 types according to their number of instructions, as shown in Table 2. Tasks are allocated randomly into 2000 VMs, which can be categorized into 4 types according to Amazon EC2 [4], as shown in Table 3. The placement of each VM is decided by the simulator according to its resource requirements. We assume batch tasks with mid-high average utilization [5] as shown in Fig. 1. To reflect the daily varying utilization, tasks are assumed to arrive according to the distribution in Fig. 2. Depending of the size of the task, its CPU utilization, and its placement, each task will run from 2 to 7 hours.

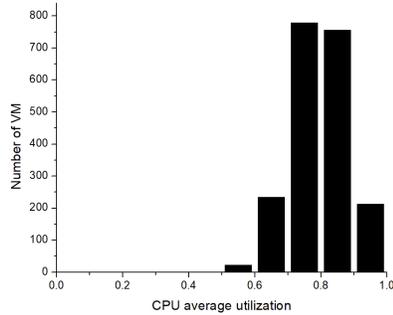


Fig. 1: CPU utilization distribution

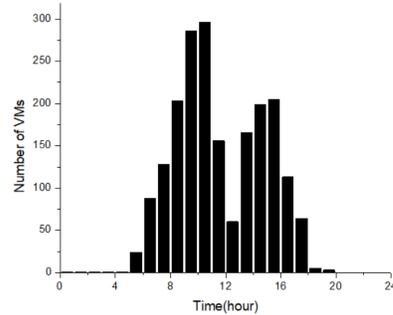


Fig. 2: Task start time distribution

5.2 Results and evaluation

Using the above settings, we simulate one day of the provider’s execution. Fig. 3 and 4 show how the cost per AU and per WU change during the experiment. The cost per AU is related to the resource allocation and the static energy consumption in the data center and the cost per WU is related to the real CPU usage from all the running VMs and their dynamic energy consumption.

We compare our model with a fixed-price model like Amazon’s [4], and a usage-based price model, such as Aldossary’s [3]. We include also an optimal price that is calculated as the cost of the dynamic energy consumed by the VM. The idea of considering only the dynamic energy cost comes from the concept of energy-proportional computing (i.e. energy should be consumed in proportion to the amount of work performed) [5]. We evaluate the total revenue for the

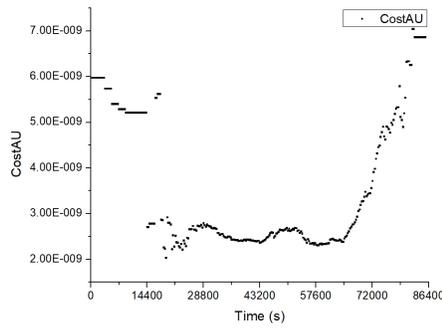


Fig. 3: CostAU changing over time

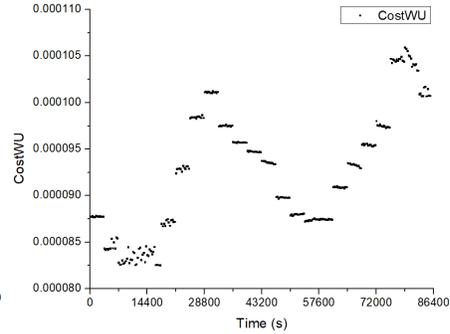


Fig. 4: CostWU changing over time

provider. We also assess the proportionality by checking how far the price of each VM is from its optimal price.

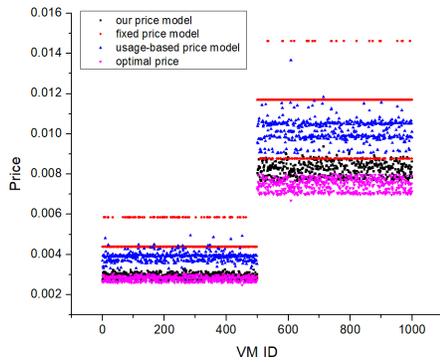


Fig. 5: Price of each VM (smaller sizes)

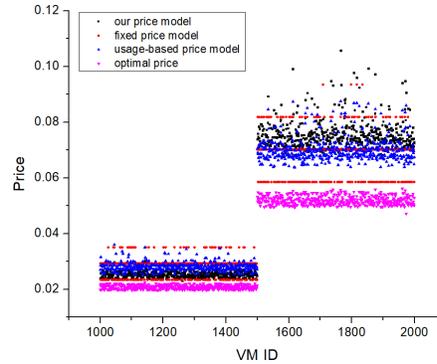


Fig. 6: Price of each VM (bigger sizes)

As shown in Table 4, all the models provide comparable revenue. Fig. 5 and 6, which display the price of each VM (ordering them by size), show that our model is the closest to the optimal for small and midsized VMs. Only big VMs pay proportionally more because they are charged for their impact in the static energy consumption. The figures confirm that the fixed model does not consider the real energy cost when pricing each VM. The usage-based model does it, but it does not allow clients to estimate their price in advance as our model does.

Table 4: Revenue comparison for the pricing models

	Our Price Model	Fixed Price Model	Usage Price Model
Revenue	56.107 €	56.113 €	56.111 €

6 Conclusions

In this paper, we have presented a pricing model for virtualized Cloud providers that is energy-aware, proportional, predictable, and fair. Our model dynamically derives the energy costs per Allocation Unit and per Work Unit for each time period. Newly arrived clients during that period can use these costs as a baseline to calculate their expenses in advance as a function of the number of requested allocation and work units. Our results demonstrate that providers can get comparable revenue to traditional pricing schemes, while offering fairer and more proportional prices to the clients than fixed-price models. Our future work will consider in the model the utilization of other resources apart from the CPU, and perform a more complex evaluation using client traces from real providers.

References

1. Make IT Green: Cloud Computing and its Contribution to Climate Change. Report, Greenpeace International (2010)
2. Aldossary, M., Djemame, K.: Energy consumption-based pricing model for cloud computing. In: 32nd UK Performance Engineering Workshop. pp. 16–27 (2016)
3. Aldossary, M., Djemame, K.: Energy-based cost model of virtual machines in a cloud environment. In: 2018 Fifth International Symposium on Innovation in Information and Communication Technology (ISIICT). pp. 1–8 (2018)
4. Amazon Web Services: Amazon EC2 Pricing, <http://aws.amazon.com/ec2/pricing/>
5. Barroso, L.A., et al.: The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines. Morgan and Claypool Publishers, 2nd edn. (2013)
6. Belady, C.L.: In the Data Center, Power and Cooling Costs More Than the IT Equipment It Supports. Electronics Cooling (February)
7. Filho, M.C.S., et al.: CloudSim Plus: A cloud computing simulation framework pursuing software engineering principles for improved modularity, extensibility and correctness. In: 2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM). pp. 400–406 (May 2017)
8. Hinz, M., et al.: A Cost Model for IaaS Clouds Based on Virtual Machine Energy Consumption. Journal of Grid Computing **16**(3), 493–512 (2018)
9. HUAWEI: Huawei server power calculator. <http://support.huawei.com/online/tools-web/ftpa/indexEn?serise=2>
10. Kansal, S., et al.: Pricing Models in Cloud Computing. In: 2014 International Conference on Information and Communication Technology for Competitive Strategies. pp. 33:1–33:5. ICTCS’14 (2014)
11. Kurpicz, M., et al.: How Much Does a VM Cost? Energy-Proportional Accounting in VM-Based Environments. In: 24th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing. pp. 651–658 (Feb 2016)
12. Microsoft Azure: Linux Virtual Machines Pricing, <https://azure.microsoft.com/en-us/pricing/details/virtual-machines/linux/>
13. Narayan, A., Rao, S.: Power-aware cloud metering. IEEE Transactions on Services Computing **7**(3), 440–451 (July 2014)
14. Rackspace: Cloud Servers Pricing and Cloud Server Costs, <http://www.rackspace.co.uk/cloud/servers/pricing>
15. Red Electrica de España: Active Energy Invoicing Price, <https://www.esios.ree.es/en/pvpc?date=17-05-2017>