Virtual Machine Power Modelling in Multi-tenant Ecosystems: Challenges and Pitfalls

Leila Sharifi[1], Felix Freitag[1], Luís Veiga[2]

[1] Computer Architecture Dept., Universitat Politecnica de Catalunya, Barcelona, Spain
[2] Técnico Lisboa/INESC-ID Lisboa, Lisboa, Portugal

Abstract— **Since energy has emerging as the first class computing resource, we need to characterize this resource in different granularity. On the other hand, the computing paradigm is shifting to the multi-tenant ecosystems. Therefore, characterizing the power consumption on Virtual Machines(VMs), running in data center hosts is necessary to attain energy efficient cloud ecosystems. In this paper, we study the challenges should be addressed in VM power modeling in cloud service provisioning.**

I. INTRODUCTION

Energy and associated environmental costs (cooling, carbon footprint, etc.) of IT services constitute a remarkable portion of service dynamic cost. Indeed, estimating energy consumption at each level of service provisioning stack, i.e. from hardware to , operating system/Virtual Machine(VM) and application is the cornerstone of research toward energy efficiency all through the stack.

Although there is a growing body of work centered on the energy aware resource management, allocation and scheduling [7,10,11] they mainly considered the whole system energy measurement, estimation, improvement and optimization. There is only limited work focusing on the energy issues per individual job [1,5,8,18]. However, they only aim at reducing total energy consumption in the infrastructure without taking into account the energy-related behavior of each individual , its performance and price, i.e., how expensive and efficient is the energy employed for the observed job performance or progress.

Nonetheless, energy-based job pricing confronts some more challenges further to the system wide energy efficiency issues. In the system wide energy efficiency, the energy consumption of the resources are measurable simply by plugging the energy meter devices or exploiting the embedded sensors of the contemporary devices. Nonetheless, it is nontrivial to measure the energy consumed per VM, since we cannot embed a physical sensor in a VM or plug a metering device to it. Therefore, estimation is still the only option in this case. Estimation results in a more complicated model since it has to deal with uncertainty and error.

In this work, we study the challenges of VM power modeling in a multi-tenant ecosystem. In the next section, we outline the background terms and hypothesis that we use in this work. Section 3 surveys the sate of the art in VM power modeling and introduces the challenges in this area. The work is concluded in Section 4.

II. BACKGROUND

In this section we introduce the background and hypothesis required to drive the discussion in the rest of this paper.

Energy Proportionality

The vision of energy proportional system implies the power model of an ideal system in which no power is used by idle systems (P_s=0), and dynamic power dissipation is linearly proportional to the system load.

LDR indicates the maximum difference of the actual power consumption, *P(U)*, and linear power model over the linear power model as in (1).

$$LDR = max \frac{P(U) - (P_s + P_d)}{P_s + P_d} \quad (1)$$

IPR is the indicator of idle to peak power consumption as illustrated in (2)

$$IPR = \frac{P_{idle}}{P_{Max}} \quad (2)$$

To measure how far a system power model is from the ideal (energy proportional) one, Proportionality Gap(PG) (19) is defined as the normalized difference of the real power value and the ideal power value, which is indicated as $P_{Max} \times U$, under a certain utilization level as shown in (3). Therefore, having proportionality gap values for a given device, we can reconstruct the power model of the device.

$$PG(U) = \frac{P(U) - (P_{Max} \times U)}{P_{Max}} \quad (3)$$

Given the state of the art hardware, designing hardware which is fully energy proportional remains an open challenge, power model of a non-energy proportional system is illustrated in Figure 1. However, even in the absence of redesigned hardware, we can approximate the behavior of energy proportional systems by leveraging combined power saving mechanisms [16] and engaging heterogeneous commodity devices combined with powerful server machines in lieu of homogeneous server hardware platform [19].

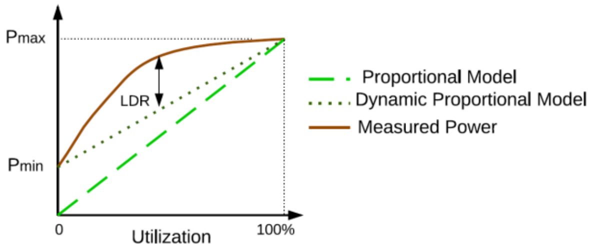

Figure 9 - Energy Proportionality.

II.1. Server power modeling

State of the art platforms are not capable of fine-grained power measurement. Therefore, to manage dynamic power proportionality, a power model is required. Currently, Running Average Power Limit (RAPL) counters, available in the recent Intel CPUs, is the closest to the hardware based monitoring. RAPL allows monitoring of whole CPU package, cores, and DRAM. Since these counters are not available in all CPUs, to cope with the heterogeneous infrastructure, a group of works rely on performance counters for synthesising a power model. Mapping the counter and power values is usually done through linear regression. Nonetheless, linear power model is not sufficient in many cases due to non-proportional power dissipation characteristics of CPUs.

Moreover, linear models rely on the non-correlated covered features, which is not a valid assumption in the state of the art systems. A quadratic solution fits better the power modeling of multi-core systems [2].

However, Hyperthreading and turbo-boost may still impede the model from accurate estimation, due to hidden states they make. A hyperthread aware power modeling mechanism is introduce in [20]. The introduced model differentiates between the cases where either single or both hardware threads of a core are in use.

The most recent work in this line is BitWatts [5], which introduces a counter based power model for each individual frequency.

Nonetheless, there is a trade off between accuracy and the overhead. Targeting the community of commodity devices forming a collaborative system, in edge layer, integrated with data center to form P2P assisted clouds, we should particularly tune the trade off level due to the lower energy consumption in such devices, which acquires adaptive models.

III. Estimating VM Energy

In multi-tenant platforms, the efficiency of VM consolidation, power dependent cost modeling, and power provisioning are highly dependent on accurate power models. Such models are particularly needed because it is not possible to attach a power meter to a virtual machine. In this section we review the state of the art VM power models and demonstrate their shortcomings.

III.1. Related Work

In general, VMs can be monitored as black-box systems for coarse-grained scheduling decisions. However, for fine-grained scheduling decisions—e.g., with heterogeneous hardware—finer-grained estimation at sub-system level is required and might even need to step inside the VM.

So far, fine-grained power estimation of VMs required profiling each application separately. To exemplify, WattApp [12], which relies on application throughput instead of performance counters as a basis for the power model. PMapper [17] maps resources using a centralized step-wise decision algorithm in lieu of application power estimation.

To generalize power estimation, JouleMeter [9] assumes that each VM only hosts a single application and thus treat VMs as black boxes. In a multi-VM system, they try to compute the resource usage of each VM in isolation and feed the resulting values in a power model. Bertran et al. [1] propose an approach employes a sampling phase to gather data related to performance-monitoring counters (PMCs) and compute energy models from these samples. With the gathered energy models, it is possible to predict the power consumption of a process, and therefore apply it to estimate the power consumption of the entire VM. Another example is VMeter [3], which estimates the consumption of all active VMs on a system. A linear model is used to compute the VMs' power consumption with the help of available statistics (processor utilization and I/O accesses) from each physical node. The total power consumption is subsequently computed by summing the VMs' consumption with the power consumed by the infrastructure.

Janacek et al. [6] exploit a linear power model to compute the server consumption with postmortem analysis. The computed power consumption is then mapped to VMs depending on their load. This technique is not effective when runtime information is required. In general, VMs can be monitored as black-box systems for coarse-grained scheduling decisions. However, for fine-grained scheduling decisions—e.g., with heterogeneous hardware— finer-grained estimation at sub-system level is required and might even need to step inside the VM.

So far, fine-grained power estimation of VMs required profiling each application separately.To exemplify, WattApp [12], which relies on application throughput instead of performance counters as a basis for the power model. PMapper [17] maps resources using a centralized step-wise decision algorithm in lieu of application power estimation.

IV. Challenges

However, collocation of applications has its own challenges. Workload intensity is often highly dynamic. The power profile of the data center hardware is inherently heterogeneous; this makes the optimal VM power modeling problem more complicated. The nonlinearity and in some cases unpredictability of the energy efficiency profile aggravates the complexity of energy efficient collocation management, due to inaccurate VM power characterization.

IV.1. Static Power and Overhead Modeling

Energy consumption of the host per job embraces the static power consumption, independent of the resource utilization, and the dynamic power, which is degraded not only proportional to the VM's allocated resources but also on account of the overhead caused in the hypervisor, and the interference due to collocation. Estimating this overhead is complicated since the pattern of the hypervisor overhead is tightly coupled with the number of VMs, the type of resources each VM asks for, and the number of times the switching occurs between VMs and hypervisor. Thus, for a more accurate estimation, further to individual VM's energy, VM interference energy overhead should also be estimated. Some estimation methods have been proposed in the state of the art: e.g. [3,4,14]. In [15] the authors argue that, in virtualized environments, energy monitoring has to be integrated within the VM as well as the hypervisor.

Work in [13] introduces an interference coefficient, defined to model the energy interference. The major contribution of this work is to estimate the energy interference according to the previous knowledge of standalone application running on the same machine. They model interference as a separate implicit task. Moreover, an energy efficient collocation management policy is introduced in this work that is modeled as an optimization problem solvable by data mining techniques. All the VMs running on the same machine are known as a collection. The energy consumption of a collection is the sum of idle energy consumed for the longest VM run, dynamic energy consumed by each VM if they were run in isolated environment, and the energy depleted due to interference between each VM pair. The interference energy can be positive or negative depending on the intersection of resources between each VM pair. Interference energy is estimated as the coefficient of the summation of idle and isolated run for each VM. On the other hand, performance is measured as the delay, which is measured by modeling the system as a M/M/1 queue and calculating the imaginary interference tasks response time as the delay due to interference.

IV.2. Non-energy Proportional Host Effect

Besides the hypervisor and interference overhead in multi-tenant systems, the non-energy proportional hardware adds more complexity to the VM power modeling agenda. In non-energy proportional hardware platform, since the hardware power model is non-linear, two identical VMs, sharing the same hardware, may end up with different dynamic power usage estimation during the runtime, which may lead to unfair energy based service charging, and planning.

Figure 2, visualizes such a case. In this scenario, there are two identical VMs, i.e. VM_1 and VM_2, collocated on a host with the power model demonstrated in the Figure.

If we only run VM_1, the dynamic power estimated for this VM will be P_1, whereas running the second identical VM on the same machine predicted as $P_2 < P_1$. Therefore, in case of collocation, there should be a strategy to divide the dynamic power fairly among the running VMs.

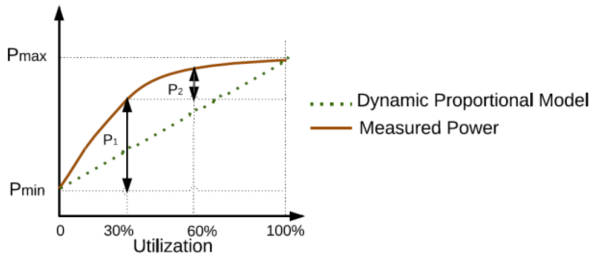

Figure 10 - VM power modeling issues in non-energy proportional systems.

To address the fairness issue introduced in the previous section we propose the weighted division VM power model. In this model as illustrated in [4], a particular VM's power consumption, $P_{VM}(i)$ is calculated according to the relative utilization, i.e. $u(i)/U$, contributed by that particular VM. In this equation, $u(i)$ represents the utilization incurred by VM i, and U denotes the overall machine utilization.

$$P_{VM}(i) = \frac{u_i P(U)}{U} \tag{4}$$

V. Conclusion

In this paper, we explained the state of the art Virtual Machine(VM) power modeling techniques and their shortcomings. We demonstrated that current VM power models, fail to capture the effect of non-energy proportional hosts in a multi-tenat cloud ecosystems. We argued that a fair VM power model needs not only to be able to characterize per VM resource usage and translate it to power, but also it requires to be aware of the overall host utilization for fair power division.

Moreover, interference and other vitalization overhead require an accurate model which can be mappable to the power consumption. Therefore, future line of work in VM power modeling should address overhead modeling in multi-tenant ecosystems baring in mind the properties of non-energy proportional hosts.

References

[65] Bertran, R.; Gonzàlez, M.; Martorell, X.; Navarro, N. & Ayguadé, E. “Counter-based power modeling methods: Top-down vs. bottom-up” The Computer Journal, Br Computer Soc, 2013, 56, 198-213.

[66] Bircher, W. L. & John, L. K. “Complete system power estimation: A trickle-down approach based on performance events Performance Analysis of Systems & Software”, 2007. ISPASS 2007. IEEE International Symposium on, 2007, 158-168.

[67] Bohra, A. E. & Chaudhary, V. VMeter: “Power modelling for virtualized clouds Parallel & Distributed Processing”, Workshops and Phd Forum (IPDPSW), 2010 IEEE International Symposium on, 2010, 1-8.

[68] Chen, Q.; Grosso, P.; van der Veldt, K.; de Laat, C.; Hofman, R. & Bal, H. “Profiling energy consumption of VMs for green cloud computing”, Dependable, Autonomic and Secure Computing (DASC), 2011 IEEE Ninth International Conference on, 2011, 768-775.

[69] Colmant, M.; Kurpicz, M.; Felber, P.; Huertas, L.; Rouvoy, R. & Sobe, A. “Process-level power estimation in VM-based systems”, European Conference on Computer Systems (EuroSys), 2015, 14.

[70] Janacek, S.; Schroder, K.; Schomaker, G.; Nebel, W.; Ruschen, M. & Pistoor, G. “Modeling and approaching a cost transparent, specific data center power consumption”, Energy Aware Computing, 2012 International Conference on, 2012, 1-6.

[71] Jin, X.; Zhang, F.; Wang, L.; Hu, S.; Zhou, B. & Liu, Z. “A Joint Optimization of Operational Cost and Performance Interference in Cloud Data Centers”, arXiv preprint arXiv:1404.2842, 2014.

[72] Kaewpuang, R.; Chaisiri, S.; Niyato, D.; Lee, B. & Wang, P. “Cooperative Virtual Machine Management in Smart Grid Environment”, IEEE, 2013.

[73] Kansal, A.; Zhao, F.; Liu, J.; Kothari, N. & Bhattacharya, A. A. ,”Virtual machine power metering and provisioning”, Proceedings of the 1st ACM symposium on Cloud computing, 2010, 39-50.

[74] Khosravi, A.; Garg, S. K. & Buyya, R. “Energy and carbon-efficient placement of virtual machines in distributed cloud data centers”, Euro-Par 2013 Parallel Processing, Springer, 2013, 317-328.

[75] Kim, N.; Cho, J. & Seo, E. “Energy-credit scheduler: An energy-aware virtual machine scheduler for cloud systems”, Future Generation Computer Systems , 2012.

[76] Koller, R.; Verma, A. & Neogi, A. “WattApp: an application aware power meter for shared data centers”, Proceedings of the 7th international conference on Autonomic computing, 2010, 31-40.

[77] Pore, M.; Abbasi, Z.; Gupta, S. K. & Varsamopoulos, G. “Energy aware Colocation of Workload in Data centers”, High Performance Computing (HiPC), 2012 19th International Conference on, 2012, 1-6.

[78] Smith, J. W.; Khajeh-Hosseini, A.; Ward, J. S. & Sommerville, I. “CloudMonitor: Profiling Power Usage”, Cloud Computing (CLOUD), 2012 IEEE 5th International Conference on, 2012, 947-948.

[79] Stoess, J.; Lang, C. & Bellosa, F. “Energy Management for Hypervisor-Based Virtual Machines”. USENIX annual technical conference, 2007, 1-14.

[80] Tolia, N.; Wang, Z.; Marwah, M.; Bash, C.; Ranganathan, P. & Zhu, X. “Delivering Energy Proportionality with Non Energy-Proportional Systems-Optimizing the Ensemble”. HotPower, 2008, 8, 2-2.

[81] A. Verma P. Ahuja, A. N. “pMapper: power and migration cost aware application placement in virtualized systems”, 9th ACM/IFIP/USENIX International Conference on Middleware , 2008, 243-264.

[82] Wang, C.; Urgaonkar, B.; Kesidis, G.; Shanbhag, U. V. & Wang, Q. “A case for virtualizing the electric utility in cloud data centers”, Proceedings of the 6th USENIX conference on Hot Topics in Cloud Computing, 2014, 22-22.

[83] Wong, D. & Annavaram, M. “Knightshift: Scaling the energy proportionality wall through server-level heterogeneity”, Microarchitecture (MICRO), 2012 45th Annual IEEE/ACM International Symposium on, 2012, 119-13.

[84] Zhai, Y.; Zhang, X.; Eranian, S.; Tang, L. & Mars, J. “Happy: Hyperthread-aware power profiling dynamically”, 2014 USENIX Annual Technical Conference (USENIX ATC 14)(Philadelphia, PA, 2014, 211-217.