A Protocol to Assess the Accuracy of Process-Level Power Models

1st Emile Cadorel
Davison consulting
Rennes, France
emile.cadorel@davidson.fr

2nd Dimitri Saingre
Davison consulting
Rennes, France
dimitri.saingre@davidson.fr

***Abstract*—The energy consumption of servers has been a critical research area, drawing significant interest from academic and industrial sectors. Various models have been developed to estimate power consumption of computing devices, ranging from simple linear models to more complex ones. In recent years, as the cloud paradigm has expanded and allowed applications to be hosted alongside others in virtual machines, power models have evolved to distribute energy usage among applications on a server. These models aim to achieve two primary objectives: monitoring the energy footprint and optimizing energy consumption. Nonetheless, little attention has been given in academic literature to evaluate the efficiency and accuracy of the allocation phase of this models. This paper presents a definition of power division and a protocol to evaluate models using such division. The proposed protocol is used to evaluate models on physical machines with different performance settings, toggling hyperthreading and turboboost. Before discussing the conceptual distinctions between power and energy allocation models, our results show the existence of some limitations in the existing models. These results provide valuable insight into the missing information needed to improve the accuracy of power models.**

***Index Terms*—energy consumption, power usage, modeling, allocation, virtual machines, performances**

I. Introduction

The energy consumption of servers is a research topic on its own that has attracted increasing interest in both the academic and industrial worlds in recent years. Many models have been developed to estimate the energy consumption of computing devices, ranging from simple linear models to more complex models. Historically such models were designed to represent the consumption of the entire computing device, (e.g. a server in an HPC cluster [1], [2], or a battery-powered mobile device [3]). In more recent years, due to the increasing use of the Cloud paradigm, applications are no longer hosted alone on a server, but are collocated with other applications, usually running inside virtual machines. To account for this new reality, power models have been adapted to divide the estimated consumption among the different applications that are running on the server [4]–[6]. In this paper, we will refer to this type of model as a power division model. Such models use sensors to acquire the energy consumption of the targeted server (e.g. RAPL), and use system metrics (e.g. performance counters) to divide this consumption among applications.

The objective generally put forward by the authors of power division model is to assign an energy consumption to the various software programs running on a given server, and to use this information to determine which software program is the most energy-consuming among the others. This objective can be analogized to the creation of *Life Cycle Assessment* (LCA) [7], [8] for the different piece of software running on a given infrastructure. Some power division models extend the system of division of energy among software applications, to division among virtual machines (VMs). In that context, VMs may belong to different users, and context of deployment (host machine, CPU architecture and size, neighbour VMs. . .) is invisible within the virtual machines. This energy consumption may be divided a second time among the applications running inside the VM. Two different actors that might be interested by power division models can be identified. a) the Cloud provider, to divide the consumption among the different VMs running on their infrastructure, and to allocate these consumptions to their clients, b) the end users, to assign the energy consumptions of their VMs to the different applications that are being executed inside their VMs.

On the other hand, it is claimed [4] that power division models can be used to optimize the energy consumption of running applications, and can help developers to understand their behaviors, in a context where multiple applications are running concurrently. From this point of view, a leverage is generally brought forward [7], [8]; modifying a preexisting source code to optimize its performance in terms of its energy consumption. It is worth noting that power division models are intended for use in a shared context where multiple applications run concurrently (this is the main reason for division) as well as in a production context. Unlike in HPC environments, where hosts are configured and optimized to accommodate specific applications, production platforms are set up to accommodate a variety of workloads, and end-users have no control over these settings. The emergence of models like Kepler [9], which target deployments in Kubernetes, demonstrate a willingness to utilize these models within a Cloud context. From our perspective, Cloud represents the primary use case for these models, and this paper will explore this context.

However, the literature has given little attention to validating the allocation algorithm utilized by power division models, claiming that there is no truth value to compare to, and have limited their investigation to the ability of the models to recover the global consumption of the machine. Our contribution provides a clear definition of power division and a standardized protocol for evaluating models utilizing such division. The proposed protocol is used to evaluate models on physical machines with different performance settings, toggling hyperthreading and turboboost. Our results reveal certain limitations in the existing power division models, which should be addressed to enhance their accuracy.

Our research work is accessible at: https://hal.science/hal-04720926.

References

[1] C. M. Olschanowsky, T. Rosing, A. Snavely, L. Carrington, M. M. Tikir, and M. Laurenzano, "Fine-grained energy consumption characterization and modeling," in *2010 DoD High Performance Computing Modernization Program Users Group Conference*, 2010.

[2] M. Jarus, A. Oleksiak, T. Piontek, and J. Weglarz, "Runtime power usage estimation of hpc servers for various classes of real-life applications," *Future Generation Computer Systems*, 2014.

[3] P. K. D. Pramanik, N. Sinhababu, B. Mukherjee, S. Padmanaban, A. Maity, B. K. Upadhyaya, J. B. Holm-Nielsen, and P. Choudhury, "Power consumption analysis, measurement, management, and issues: A state-of-the-art review of smartphone battery and energy usage," *IEEE Access*, 2019.

[4] G. Fieni, R. Rouvoy, and L. Seinturier, "SmartWatts: Self-Calibrating Software-Defined Power Meter for Containers," in *CCGRID 2020*, Melbourne, Australia, 2020.

[5] M. Colmant, R. Rouvoy, M. Kurpicz, A. Sobe, P. Felber, and L. Seinturier, "The next 700 cpu power models," *Journal of Systems and Software*, 2018.

[6] M. Jay, V. Ostapenco, L. Lefèvre, D. Trystram, A.-C. Orgerie, and B. Fichel, "An experimental comparison of software-based power meters: focus on CPU and GPU," in *CCGrid 2023*, Bangalore, India, 2023.

[7] S. Georgiou, S. Rizou, and D. Spinellis, "Software development lifecycle for energy efficiency: Techniques and tools," *ACM Comput. Surv.*, 2019.

[8] N. Rashid and S. U. Khan, "Agile practices for global software development vendors in the development of green and sustainable software," *Journal of Software: Evolution and Process*, 2018.

[9] [Online]. Available: https://sustainable-computing.io/