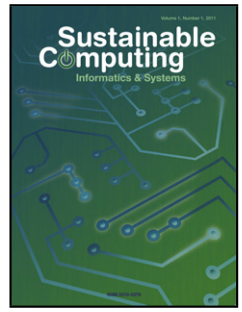
Accepted Manuscript



Title: Scheduling-based Power Capping in High Performance Computing Systems

Author: Andrea Borghesi Andrea Bartolini Michele Lombardi Michela Milano Luca Benini

PII:	S2210-5379(17)30231-7
DOI:	https://doi.org/doi:10.1016/j.suscom.2018.05.007
Reference:	SUSCOM 247

To appear in:

Received date:	3-7-2017
Revised date:	27-11-2017
Accepted date:	15-5-2018

Please cite this article as: Andrea Borghesi, Andrea Bartolini, Michele Lombardi, Michela Milano, Luca Benini, Scheduling-based Power Capping in High Performance Computing Systems, *<![CDATA[Sustainable Computing: Informatics and Systems]]>* (2018), https://doi.org/10.1016/j.suscom.2018.05.007

Scheduling-based Power Capping in High Performance Computing Systems

Andrea Borghesi[a,b], Andrea Bartolini[b,c], Michele Lombardi[a], Michela Milano[a], Luca Benini[b,c]

[a]*DISI, University of Bologna. Viale Risorgimento 2, 40123, Bologna, Italy*
[b]*DEI, University of Bologna. Viale Risorgimento 2, 40123, Bologna, Italy*
[c]*Integrated Systems Laboratory at ETH Zurich, Switzerland*

Abstract

Supercomputer installed capacity worldwide increased for many years and further growth is expected in the future. The next goal for High Performance Computing (HPC) systems is reaching Exascale. The increase in computational power threatens to lead to unacceptable power demands, if future machines will be built using current technology. Therefore reducing supercomputer power consumption has been the subject of intense research. A common approach to curtail the excessive power demands of supercomputers is to hard-bound their consumption, *power capping*. Power capping can be enforced by reactively throttling system performance when the power bound is hit, or by scheduling workload in a proactive fashion to avoid hitting the bound. In this paper we explore the second approach: our scheduler meets power capping constraints and minimizes Quality-of-Service (QoS) disruption through smart planning of the job execution order. The approach is based on Constraint Programming in conjunction with a Machine Learning module predicting the power consumptions of HPC applications. We evaluate our method on the Eurora supercomputer, using both synthetic workloads and historical traces. Our approach outperforms the state-of-the-art power capping techniques in terms of waiting time and QoS, while keeping schedule computation time under control.

Keywords: Constraint Programming, Optimization, HPC, Power Consumption, Scheduling, Machine Learning, Power Modeling

1. Introduction

Power consumption is a critical limiter for next generation High Performance Computing systems: supercomputers are expected to reach Exascale in 2023 [1], as revealed by the increase of the worldwide supercomputer installation [2], but at the price of unsustainable power demand growth. Today's most powerful system is Sunway TaihuLight which reaches 93 PetaFlops with 15.371 MWatts of power dissipation [3]. An Exascale machine built with current technology would consume an excessive amount of power (hundreds of MWatts), while a commonly accepted upper bound for a supercomputer power consumption is around 20MW [4]. Therefore, in the last years the HPC community put great effort in finding effective ways to reduce power consumption of HPC facilities, either developing new hardware and software solutions or optimizing the management of existing systems.

Many strategies try to limit the power consumption within a certain power budget, never to be exceeded. These methods are generally referred to as *power capping* [5]. A big challenge for the adoption of power capping solutions is the need to find a good balance between curtailing power consumption and keeping a high level of Quality-Of-Service for the system users. In particular, most power capping approaches rely on decreasing the performance of the computing nodes (i.e. imposing a node power budget and/or decreasing the operational frequency as a reactive countermeasure when the power cap is reached) which in turn leads to increased duration of HPC jobs or longer waiting times. Increased durations can produce user dissatisfaction due to longer completion times and possibly increased costs. In fact the accounting policy in some current supercomputers links the price for the user to the duration (number of CPU-hours) of the job.

We reckon that a key role in this challenge can be played by the scheduling software that decides where and when a job has to execute, a software module commonly referred to as *job dispatcher*. We claim that with a "clever" job dispatcher it is possible to operate a power capped system at a higher Quality-of-Service than using reactive performance throttling techniques that are in common use today. Constraint Programming (CP) is a paradigm to solve NP-hard problems by exploring a set of feasible solutions optimizing one or multiple objective functions. However, this technique is not widely used in HPC facilities, because the run-time of solvers are not compatible with the online nature of job schedulers for supercomputers. On the other hand, supercomputer jobs do have a significant duration and their arrival rate is significantly lower than that of e.g. data centers and enterprise servers workload. This creates opportunity for an optimization-based scheduling. The feasibility of this approach has been already demonstrated by several works [6, 7].

In this paper we propose a job dispatcher able to limit a supercomputer power consumption acting only on the workload scheduling. Our approach satisfies the power constraint work-

Email addresses: andrea.borghesi3@unibo.it (Andrea Borghesi), a.bartolini@unibo.it, barandre@iis.ee.ethz.ch (Andrea Bartolini), michele.lombardi2@unibo.it (Michele Lombardi), michela.milano@unibo.it (Michela Milano), luca.benini@unibo.it,luca.benini@iis.ee.ethz.ch (Luca Benini)