Power Capping in High Performance Computing Systems

Andrea Borghesi\textsuperscript{1,2}, Francesca Collina\textsuperscript{1,2}, Michele Lombardi\textsuperscript{1,2}, Michela Milano\textsuperscript{1,2}, and Luca Benini\textsuperscript{1,2}

\textsuperscript{1} DISI, University of Bologna, Bologna, Italy
\textsuperscript{2} DEI, University of Bologna, Bologna, Italy
{andrea.borghesi3,michele.lombardi2,michela.milano,luca.benini}@unibo.it, francesca.collina@gmail.com

Abstract. Power consumption is a key factor in modern ICT infrastructure, especially in the expanding world of High Performance Computing, Cloud Computing and Big Data. Such consumption is bound to become an even greater issue as supercomputers are envisioned to enter the Exascale by 2020, granted that they obtain an order of magnitude energy efficiency gain. An important component in many strategies devised to decrease energy usage is “power capping”, i.e., the possibility to constrain the system power consumption within certain power budget. In this paper we propose two novel approaches for power capped workload dispatching and we demonstrate them on a real-life high-performance machine: the Eurora supercomputer hosted at CINECA computing center in Bologna. Power capping is a feature not included in the commercial Portable Batch System (PBS) dispatcher currently in use on Eurora. The first method is based on a heuristic technique while the second one relies on a hybrid strategy which combines a CP and a heuristic approach. Both systems are evaluated and compared on simulated job traces.

1 Introduction

Supercomputer peak performance is expected to reach the ExaFLOP level in 2018-2020 \cite{14,15}, however energy efficiency is a key challenge to be addressed to reach this milestone. Today’s most powerful Supercomputer is Tianhe-2 which reaches 33.2 PetaFlops with 17.8 MWatts of power dissipation \cite{13}. Exascale supercomputers built upon today’s technology would lead to an unsustainable power demand (hundreds of MWatts) while according to \cite{9} an acceptable range for an Exascale supercomputer is 20MWatts; for this goal, current supercomputer systems must obtain significantly higher energy efficiency, with a limit of 50GFlops/W. Today’s most efficient supercomputer achieves 5.2 GFlops/W, thus we still need to close an order of magnitude gap to fulfill the Exascale requirements.

Almost all the power consumed by HPC systems is converted into heat. In addition to the power strictly needed for the computation - which measures only the computational efficiency - the cooling infrastructure must be taken into...
account, with its additional power consumption. The extra infrastructure needed for cooling down the HPC systems has been proved to be a decisively limiting factor for the energy performance [3]; a common approach taken to address this problem is the shift from air cooling to the more efficient liquid cooling [10].

Hardware heterogeneity as well as dynamic power management have started to be investigated to reduce the energy consumption [16][2]. These low-power techniques have been derived from the embedded system domain where they have proven their effectiveness [1]. However, a supercomputer is different from a mobile handset or a desktop machine. It has a different scale, it cannot be decoupled by the cooling infrastructures and its usage mode is peculiar: it is composed by a set of scientific computing applications which run on different datasets with a predicted end time [7]. Finally supercomputers are expensive (6 orders of magnitude more than an embedded device [19]) making it impossible for researchers to have them on their desk. These features have limited the development of ad-hoc power management solutions.

Current supercomputers cooling infrastructures are designed to withstand power consumption at the peak performance point. However, the typical supercomputer workload is far below the 100% resource utilization and also the jobs submitted by different users are subject to different computational requirements[29]. Hence, cooling infrastructures are often over-designed. To reduce overheads induced by cooling over-provisioning several works suggest to optimize job dispatching (resource allocation plus scheduling) exploiting non-uniformity in thermal and power evolutions [8][20][21][23]. Currently most of these works are based on simulations and model assumptions which unfortunately are not mature enough to be implemented on HPC systems in production, yet.

With the goal of increasing the energy efficiency, modern supercomputers adopt complex and hybrid cooling solutions which try to limit the active cooling (chiller, air conditioner) by allowing direct heat exchange with the ambient (Free-cooling). Authors in [12] show for the 2013’s top GEEN500 supercomputer that the cooling costs increase four times when the ambient temperature moves from 10 C to 40 C. Moreover the authors show that for a given ambient temperature there is a well-defined maximum power budget which guarantees efficient cooling. With an ambient temperature of 10 C the power budget which maximizes the efficiency is 45KWatt while at 40 C is 15KWatt. Unfortunately, today’s ITs infrastructure can control its power consumption only reducing the power and performance of each computing node. This approach is not suitable for HPC system where users require to execute their application in a portion of the machine with guaranteed performance. A solution commonly adopted in HPC systems is power capping [25][17][15][27][22], which means forcing a supercomputer not to consume more than a certain amount of power at any given time. In this paper we study a technique to achieve a power capping by acting on the number of job entering the system.