Sustainable Computing: Informatics and Systems 3 (2013) 183–193

Contents lists available at SciVerse ScienceDirect

Sustainable Computing: Informatics and Systems

journal homepage: www.elsevier.com/locate/suscom

The need for speed and stability in data center power capping

Arka A. Bhattacharya[a,*], David Culler[a], Aman Kansal[b], Sriram Govindan[c], Sriram Sankar[c]

[a] University of California, Berkeley, Berkeley, CA, USA
[b] Microsoft Research, Redmond, WA, USA
[c] Microsoft Corporation, Redmond, WA, USA

ARTICLE INFO

Article history:
Received 11 September 2012
Accepted 29 January 2013

Keywords:
Power capping
Admission control
Frequency scaling

ABSTRACT

Data centers can lower costs significantly by provisioning expensive electrical equipment (such as UPS, diesel generators, and cooling capacity) for the actual peak power consumption rather than server nameplate power ratings. However, it is possible that this under-provisioned power level is exceeded due to software behaviors on rare occasions and could cause the entire data center infrastructure to breach the safety limits. A mechanism to *cap* servers to stay within the provisioned budget is needed, and processor frequency scaling based power capping methods are readily available for this purpose. We show that existing methods, when applied across a large number of servers, are not fast enough to operate correctly under rapid power dynamics observed in data centers. We also show that existing methods when applied to an open system (where demand is independent of service rate) can cause cascading failures in the software service hosted, causing the service performance to fall uncontrollably even when power capping is applied for only a small reduction in power consumption. We discuss the causes for both these short-comings and point out techniques that can yield a safe, fast, and stable power capping solution. Our techniques use admission control to limit power consumption and ensure stability, resulting in orders of magnitude improvement in performance. We also discuss why admission control cannot replace existing power capping methods but must be combined with them.

1. Introduction

The cost of provisioning power in data centers is a very large fraction of the total cost of operating a data center [1–3] ranking just next to the cost of the servers themselves. *Provisioning* costs include the cost of infrastructure for sourcing, distribution and backup for the peak power capacity (measured in $/kW). These are higher than the *consumption* costs paid per unit of energy actually consumed (measured in $/kWh) over the life of a data center. Provisioned capacity and related costs can be reduced by minimizing the peak power drawn by the data center. A lower capacity saves on expenses in utility connection charges, diesel generators, backup batteries, and power distribution infrastructure within the data center. Lowering capacity demands is also greener because from the power generation standpoint, the cost and environmental impact for large scale power generation plants such as hydro-electric plants as well as green energy installations such as solar or wind farms, is dominated by the capacity of the plant rather than the actual energy produced. From the utility company perspective, providing peak capacity is expensive due to the operation of 'peaker power plants' which are significantly more expensive to operate and are less environmentally friendly than the base plants. Aside from costs, capacity is now is short supply in dense urban areas, and utilities have started refusing to issue connections to new data centers located in such regions. Reducing the peak power capacity required is hence extremely important.

The need to manage peak power is well understood and most servers ship with mechanisms for power capping [4,5] that allow limiting the peak consumption to a set threshold. Further capacity waste can be avoided by coordinating the caps across multiple servers. For instance, when servers in one cluster or application are running at lower load, the power left unused could be used by other servers to operate at high power levels than would be allowed by their static cap. Rather than forcing a lower aggregate power level at all times, methods that coordinate the power caps dynamically across multiple servers and applications have been developed [6–10].

We identify two reasons why existing power capping methods do not adequately meet the challenge of power capping in data centers. The first is *speed*. We show through real world data center power traces that power demand can change at a rate that is too fast for the existing methods. The second is *stability*. We experimentally show that when hosting online applications, the

[*] Corresponding author.
E-mail addresses: arka@eecs.berkeley.edu, arkaaloke@gmail.com (A.A. Bhattacharya), culler@eecs.berkeley.edu (D. Culler), kansal@microsoft.com (A. Kansal), srgovin@microsoft.com (S. Govindan), sriram.sankar@microsoft.com (S. Sankar).

2210-5379/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.suscom.2013.01.005

system may become unstable if power capped. A small reduction in power achieved through existing power capping methods can cause the application latency to increase uncontrollably and may even reduce throughput to zero. We focus on the importance of the two necessary properties – *speed* and *stability*, and propose ways of achieving them and discuss the tradeoffs involved. Our observations are generic, and can be integrated into any power capping algorithm.

Specifically, the paper makes the following contributions:

- We quantify the benefit of using power capping to lower power provisioning costs in data centers through the analysis of a real world data center power trace.
- *Speed requirement:* From the same trace, we characterize the rates at which power changes in a data center. We make a case for one-step power controllers by showing that existing closed-loop techniques for coordinated power capping across a large number of servers may not be fast enough to handle data center power dynamics.
- *Stability requirement:* We show that existing power capping techniques do not explicitly shape demand, and can lead to instability and unexpected failures in online applications.
- We present admission control as a power capping knob. We demonstrate that admission control integrated with existing power capping techniques can achieve desirable stability characteristics, and evaluate the trade-offs involved.

2. Power costs and capping potential

Most new servers ship with power capping mechanisms. System management software, such as Windows Power Budgeting Infrastructure, IBM Systems Director Active Energy Manager, HP Insight Control Power Management v.2.0, Intel Node Manager, and Dell OpenManage Server Administrator, provide APIs and utilities to take advantage of the capping mechanisms. In this section we discuss why power capping has become a significant feature for data centers.

2.1. Power provisioning costs

The designed peak power consumption of a data center impacts both the capital expense of provisioning that capacity as well as the operating expense of paying for the peak since there is often a charge for peak usage in addition to that for energy consumed.

The capital expense (cap-ex) includes power distribution infrastructure as well as the cooling infrastructure to pump out the heat generated from that power, both of which depend directly on the peak capacity provisioned. The cap-ex varies from $10 to $25 per Watt of power provisioned [3]. For example, a 10 MW data center spends about $100–250 million in power and cooling infrastructure. Since the power infrastructure lasts longer than the servers, in order to compare this cost as a fraction of the data center expense, we can normalize all costs over the respective lifespans. Amortizing cap-ex over the life of the data center (12–15 years [3,2]), server costs over the typical server refresh cycles (3–4 years), and other operating expenses at the rates paid, the cap-ex is *over a third* of the overall data center expenses [11,2]. This huge cost is primarily attributable to the expensive high-wattage electrical equipment, such as UPS batteries, diesel generators, and transformers, and is further exacerbated by the redundancy requirement mandated by data center availability stipulations.

The peak power use affects operating expenses (op-ex) as well. In addition to paying a per unit energy cost (typically quoted in $/kWh), there is an additional fee for the peak capacity drawn, even if that peak is used extremely rarely. Based on current utility tariffs [12] for both average and peak power, the peak consumption can contribute to as much as 40% of the utility bill [13]. Utility companies may also impose severe financial penalties for exceeding contracted peak power limits.

The key implication is that reducing the peak capacity required for a data center, and adhering to it, is highly beneficial.

2.2. Lower cost through capping

Power capping can help manage peak power capacity in several ways. We describe some of the most common reasons to use it below.

2.2.1. Provisioning lower than observed peak

Probably the most widely deployed use case for power capping is to ensure safety when power is provisioned for the actual data center power consumption rather than based on server *nameplate ratings*. *Nameplate ratings* on servers denotes its maximum possible power consumption, computed as the sum of maximum power consumption of all the server sub-components and a conservative safety margin. The name-plate rating on servers is typically much higher than the server's actual consumption. Since no workload actually exercises every server subcomponent at its peak rated power, the name plate power is not reached in practice. Data center designers thus provision for the *observed peak* on every server. The observed peak is the maximum power consumption measured on a server when running the hosted application at the highest request rate supported by the server. This observed peak can be exceeded after deployment due to software changes or events such as server reboots that may consume more than the previously measured peak power. Server level power caps can be used to ensure that the provisioned capacity is never exceeded and protect the circuits and power distribution equipment.

Server level caps do not eliminate waste completely. Setting the cap at each server to its observed peak requires provisioning the data center for the *sum of the peaks*, results in wasted capacity since not all servers operate at the peak simultaneously. Instead, it is more efficient to provision for the *peak of the sum* of server power consumptions, or equivalently, the estimated peak power usage of the entire data center. The estimate is based on previously measured data and may sometimes be exceeded. Thus a cap must be enforced at the data center level. Here, the server level caps will change dynamically with workloads. For instance, a server consuming a large amount of power need not be capped when some other server has left its power unused. However the former server may have to be capped when the other server starts using its fair share. Coordinated power capping systems [6–10] can be used for this.

Additionally, even the observed peak is only reached rarely. To avoid provisioning for capacity that will be left unused most of the time, data centers may provision for the 99th percentile of the peak power. Capping would be required for 1% of the time, which may be an acceptable hit on performance in relation to cost savings. If the difference in magnitude of power consumed at the peak and 99th percentile is high, the savings can be significant. To quantify these savings, we present power consumption data from a section comprising of several thousand servers in one of Microsoft's commercial data centers that host online applications serving millions of users, including indexing and email workloads. The solid line in Fig. 1 shows the distribution of power usage, normalized with respect to the peak consumption. If the 99th percentile of the observed peak is provisioned for, the savings in power capacity can be over 10% of the data center peak. Capacity reduction directly maps to cost reductions.

Trends in server technology indicate that the margin for savings will increase further. Power characteristics of newer servers accentuate the difference between the peak and typical power