ECE697 GC Paper Review:

Power Provisioning for a Warehouse-sized Computer

Zhefan Lin

1. **Summary**

This paper presents the aggregate power usage characteristics of large collections of servers (up to 15 thousand) for different classes of applications over a period of approximately six months. Those observations allow authors to evaluate possibility for maximizing the use of the deployed power capacity of datacenters, and assess the risks of over-subscribing it. It found that even in well-tuned applications there is a noticeable gap (7 - 16%) between achieved and theoretical aggregate peak power usage at the cluster level (thousands of servers). The gap grows to almost 40% in whole datacenters. This headroom can be used to deploy additional compute equipment within the same power budget with minimal risk of exceeding it. In this paper, author uses their modeling framework to estimate the potential of power management schemes to reduce peak power and energy usage. They find that the opportunities for power and energy savings are significant, but greater at the cluster-level (thousands of servers) than at the rack-level (tens). They argue that systems need to be power efficient across the activity range, and not only at peak performance levels.

This paper indicates how power usage varies over time, and as the number of machines increases from individual racks to clusters of up to five thousand servers. By using multiple production workloads, the quantization in this study shows how power usage patterns are affected by workload choice. The understanding of power usage dynamics mean the optimization of power management and provisioning policies, as well as quantify the potential impact of power and energy reduction.

This paper also argues that nameplate ratings are of little use in power provision in datacenter, as they tend to grossly overestimate actual maximum usage. Using a more realistic peak power definition from this paper, it was able to quantify the gaps between maximum achieved and maximum theoretical power consumption of groups of machines. These gaps would allow hosting between 7% and 16% more computing equipment for individual (well-tuned) applications, and as much as 39% in a real datacenter running a mix of applications, through careful over-subscription of the datacenter power budget. It proves that power-capping mechanisms can act as a safety net against the risks of over-subscription, and provide additional albeit modest power savings.

1. **Strength**

This is the first power usage study at the scale of datacenter workloads, and the first reported use of model-based power monitoring techniques for power provisioning in real production systems. It has following contribution:

1. This paper points out that the gap between the maximum power actually used by large groups of machines and their aggregate theoretical peak usage can be as large as 40% in datacenters. Therefore it suggests a significant possibility to host additional machines under the same power budget in a datacenter. For well-tuned large workloads, it proves that this gap is smaller but still significant.
2. It proposes that power capping using dynamic power management can enable additional machines to be hosted, but is more useful as a safety mechanism to prevent overload situations.
3. Time intervals are observed when large groups of machines are operating near peak power levels. It suggests that power gaps and power management techniques might be more easily exploited at the datacenter-level than at the rack-level.
4. This paper gives a approach of CPU voltage/frequency scaling, a technique targeted at energy management, has the potential to be moderately effective at reducing peak power consumption at datacenter cluster level.
5. The benefits of building systems are evaluated. Power efficiency is across the activity range, instead of simply at peak power or performance levels.
6. **Weakness**

The weakness of this paper is as follows:

1. It only shows the result of estimation server power user, but it does not show the estimation model in details.
2. The data used in simulation is gathered in specific Google datacenter, it does not show any possible different simulation result based on usage of different company.
3. It proposes saving strategies based on simulation result, but it does not apply these strategies in practical datacenter and test its reliability.
4. It does not provide the infrastructure of designed datacenter with proposed saving strategies. Also, for the current running datacenters, they also need a improved plan to maximum their utilization, but this paper does not show it.
5. This paper only shows the improvement when single strategy is applied. We are not clear the consequence after we combine Mixing, CPU Voltage/Frequency Scaling and Power capping together.