Using Carbon-awareness to Reduce the Cloud’s Environmental Impact

Chaojie Zhang, Andrew A. Chien∗
University of Chicago, *Argonne National Lab
{chaojie,achien}@cs.uchicago.edu

Abstract
Carbon emissions and power are a challenge for the growth of cloud computing. We consider a carbon-aware approach to acquiring power that caps emissions hourly, adapting to the carbon content of the power grid. The objective of this approach is to reduce the cloud’s carbon footprint.

Because this approach produces variable capacity, we assess performance on production cloud workloads (Microsoft, Google), exploring several dimensions of variation, dynamic range, structure, and change frequency. The robust high performance shows that carbon-aware policies need not damage cloud efficiency. With foresight and scheduler improvement, potential job impact is minimized under extreme dynamic range.

To demonstrate the potential benefits, we consider a case study with a cloud datacenter in Germany. Our results show that the constant hourly emission approach can reduce carbon footprint by up to 19% while maintaining system performance. An additional benefit is lower power cost.

CCS Concepts: • Hardware → Impact on the environment; Enterprise level and data centers power issues; • Computer systems organization → Cloud computing.

Keywords: Resource variability, Data center, Cloud computing, Job scheduling

1 Introduction
With the end of Dennard scaling and explosive growth of demand for internet services and cloud computing, the power requirements for cloud datacenters have been growing rapidly (25-30%/year). It is no exaggeration to say that power is an important concern for large-scale cloud datacenters. And, with power usage effectiveness levels approaching 1.1, these power limits translate directly into limits on the amount of cloud computing that can be delivered. Compounding this, applications of computing drawn from every corner of commerce, society, science, and government [16, 61] are proliferating. Recent years have seen a dramatic rise in concerns about climate change, and this escalating crisis has shone a bright spotlight the cloud’s power consumption and its growing environmental impact [21, 27, 34, 40].

The scale of cloud computing infrastructure, and its rapid growth reflect rapid commercial growth, expanding at a rate of 20% per year to estimated revenue of $331 billion in 2022[6]. Along with revenue growth, for hyperscale cloud provider’s corollary growth in power consumption, rising from 10TWh in 2010 6.5x to 65 TWh in 2018 [40]. Based on a compound annual growth rate of 30%, this article projects a potential rise to over 100 TWh in just the next few years. This growth can be seen in increasing numbers of datacenters all over the world, and the drive of cloud datacenter construction causing increased power grid buildout around the world. Within the datacenters, sophisticated power management systems optimize how the power is used, according to metrics such as power usage effectiveness (PUE) [11]. Overall, hyperscale cloud datacenters have become the fastest growing consumer of electric power in many parts of the world.

Governments around the world are enacting policies to reduce carbon emissions, increasing pressure to minimize datacenter power consumption and encouraging the use of renewable generation. For example, European Union has targeted dramatic reduction of carbon emissions for the entire economy – 40% by 2030 and zero net by 2050 [14]. While the United States has no national policy, large states have adopted zero net carbon emissions goals for electric energy (California 2045) and for the entire economy (New York 2045) [15, 41, 43]. Such targets pose contrast with rapid hyperscale cloud power growth (e.g. Amazon, Microsoft, Google, etc.) and its further acceleration by machine learning [31, 52]. In Northern Virginia, datacenters account for over 5% of the power load [48], and are growing at a 20 to 40% CAGR. Several cloud computing providers have responded with aggressive goals to reduce carbon emissions – notably CEO’s Satya Nadella of Microsoft, and Sundar Pichai of Google, and even Jeff Bezos of Amazon [8–10], but there is much work to be done in the face of rapid growth of the cloud.

Of course the importance of power as a key limit has spawned a large and vibrant body of research on power capping [36, 42, 49, 62], exploitation of renewables [20, 25, 26, 32, 45] and even more radical approaches that exploit excess grid renewable energy [35, 59, 61]. These approaches reflect power’s critical relationship to computing capacity. With recent cloud systems connecting power management flexibly over clusters or even datacenters [37, 49, 58], and variation for renewables [2, 46, 61], there is an opportunity for a new approach.
We propose a new approach to reduce the carbon emissions of cloud datacenters that increases their capacity at constant emissions, or at constant capacity reduces carbon emissions. The idea is to purchase power at constant carbon emissions, which weights power acquisition towards low-carbon power, while holding total carbon emissions constant. The increased quantity of power is an opportunity, but comes with the side-effect of variable power or datacenter capacity. This scenario is depicted in Figure 1. To the cluster or datacenter resource manager, the varying power appears as variable capacity, where the available capacity of the computing resources to be scheduled changes over time (see Figure 2).

To evaluate our idea, we study how much power can be obtained with this approach, studying variation of carbon emissions. And, because current resource management systems and schedulers generally assume full knowledge of resource capacity and presume that it is stable going forward. We also study the implications of such dynamic resource capacity change on scheduler performance, to assess if the power obtained can be effectively by cloud resource managers. We study commercial cloud workloads using oversubscription resource scheduling models. First, we examine performance – goodput and SLO violations under several dimensions of capacity variation (dynamic range, variability structure, change frequency). Second, we examine techniques that might improve performance such as adding scheduler foresight and selective slowdown. Finally we consider a case study - a datacenter located in the German power grid. This specific scenario enables us to assess the realizable increases in power and reduced carbon emissions in a practical setting of real dynamic power grid variation.

Specific contributions of the paper include:

- Understanding of how carbon-emission based power acquisition can increase available power while holding carbon emissions constant;
- Study of commercial cloud workloads shows scheduler performance is robust in the face of many dimensions of capacity variation (dynamic range, change frequency, and more) of capacity variation. Of these, only extreme dynamic range significantly increases job slowdowns.
- Study of scheduler improvements shows that both capacity foresight and Long Jobs heuristic are effective to reduce most of the job slowdowns while maintaining goodput.
- A case study of a cloud datacenter in the German Power Market that shows carbon-emission based power purchase can decrease carbon emissions by up to 19% (at constant delivered capacity) or increase goodput by 3% at constant emissions.

Overall, we study how dynamism in resource capacity affects resource management efficiency in a variety of scenarios. These results highlight large potential impacts, and thereby the variability dimensions to avoid. But the results also indicate the opportunity to create new abilities to deliver goodput in the face of increasing dynamic variation.

The rest of the paper is organized as follows. We cover background in Section 2. In Section 3 we propose carbon-aware power acquisition and describe our approach. In Section 4, we introduce the system model, workload, and metrics for simulation. Section 5 presents our simulation results and analysis in understanding how resource volatility impacts resource management. We consider enhanced scheduling of foresight and job slowdown policy in Section 6. In Section 7, we conduct a case study which shows the carbon and power benefits of carbon-emission-aware power acquisition approach. We discuss related work and summarize in Section 8 and 9.

2 Background

2.1 Cloud Power Use and Datacenter Power Management

This year, 2020, has seen escalating concerns about climate change, including unprecedented California, Amazon, and Siberian wildfires, increased hurricane activities, and increased global surface temperatures. These escalations build on several decades of steadily increasing concerns that led to the Kyoto and Paris Accords. Against this backdrop, the rapid growth of the hyperscale cloud provider power consumption is striking. These organizations are documented as consuming power rising from 10TWh in 2010 6.5x to 65 TWh in 2018 [40]. This article projects a potential rise to over 100 TWh in just the next few years. Such growth is evidenced
by increasing numbers of datacenters globally. Overall, hyperscale cloud datacenters have become the fastest growing consumer of electric power in many parts of the world.

Concern about the carbon footprint has led to significant public scrutiny from organizations such as Greenpeace [5, 28]. In late 2018, Google raised the bar, adopting a goal beyond offsetting and adopting a goal of matching its power consumption on an hourly basis, 7x24 over the entire year, with renewable energy in the same power grid [27]. The sensitivity that cloud provider CEOs feel about is evidenced by their increasing outspokenness about the need to find ways to reduce the environmental damage they are causing [8–10]. This is more than noble talk; rather it is a critical self-interest to secure their opportunity for future growth.

2.2 Renewable Energy and Datacenter Adaptive Loads

With ambitious goals to de-carbonize electric power generation in much of the world, power companies and grids have turned heavily to renewable sources such as solar and wind [29, 44]. Given their volatility, these resources are often characterized by a capacity factor such as 0.33—the fraction of the nominal maximum generation that they provide over a full-year. The correlation and non-dispatchability of wind and solar generation result in diminishing benefits, “grid effective capacity factor” that diminishes with each additional unit of renewables added to the grid. This phenomena is well documented and reflects a major challenge to high-renewable fraction grids [55]. And, when renewable generation coincides, supply can exceed demand, producing negative pricing and power curtailment (waste) in massive quantities [13, 18, 29, 30]. A critical solution is adaptive loads, that adjust their demand rapidly to match the available supply [23]. Such loads will be a staple of the future grid, because of their cost-effectiveness relative to energy storage. Dynamic power management of hyperscale datacenters is an important potential adaptive load.

2.3 Resource Management

Resource management and job scheduling monitors and control resource usage, mapping jobs onto a set of machines, optimizing metrics such as makespan, job wait time, goodput, and resource utilization. While existing data centers deal with a great variety of workloads, such as streaming and interactive jobs, batch workloads are an important workload.

While traditional schedulers dedicate resources to jobs based on their requests, cloud computing systems generally use oversubscription. That is, allocating a resource to multiple jobs, and depending on their statistical variability to enable them to co-exist and achieve expected performance. It is the fundamental mechanism that enables cloud data centers to substantially improve efficiency by over-committing resources multiple times. This approach achieves much greater loading of computing hardware – and thus greater efficiencies or revenues.

Oversubscription exploits the fact that many jobs exhibit low average resource utilization – far less than requested [19, 39, 54]. Oversubscription schedulers can improve system throughput and resource utilization, as well as low latency for production jobs through statistical multiplexing of workloads. Of course, the level of oversubscription has to be carefully designed and tuned in case of unexpected spikes in usage [19]. These designs include but are not limited to complex characterization and prediction of resource utilization, cluster deployment size, server maintenance, and appropriate allocation size.

3 The Opportunity of Carbon-emission Variation

Modern power grids include a complex mix of generators—wind, solar, hydro, as well as fossil-fuel and even nuclear. As load varies through the day or over the week, the power grid dynamically dispatches generators in an ever-changing mix to meet the current demand. While generally preference is given to renewables through economic dispatch because they have low incremental generation cost, they are not always available in sufficient quantity so carbon-emitting generators are used. This problem is much harder than most markets because power is what economists call a perishable resource—generation much be matched instantaneously with load.

The net effect is that carbon-emissions content of power in most power grids varies widely with time, producing profiles such as shown in Figure 2 when power is purchased based on a constant hourly carbon-emissions budget. Resulting variable power ranges from 16 to 42MW, comparing to a constant power of 28MW. Not that only does the level of power vary widely, the carbon-emissions based purchases

![Figure 2. Resource capacity variation for a datacenter with a fixed per-hour carbon budget](image)
also exploit this variation to buy a significantly larger quantity of power at the same level of carbon emissions. In Figure 2(b), more than 10% capacity increase is observed over mai. The example illustrates that following carbon-emission content of power market alone can produce resource capacities ranging from 40% to 105% within a single 24-hour period. As renewable generation increases, carbon content in power grids will continue to exhibit increasingly larger differences.

When resource capacity varies it affects resource management. The magnitude of capacity change can be large; the example shows a factor of two between the low and high. So, how useful is the variable capacity? Traditional schedulers assume constant resource capacity. Based on the assumption that current capacity will continue, these schedulers make decisions that commit resources into the future. Because they have been designed to maximize goodput, they strive to fill as much of this capacity as possible. If resource capacity decreases, the schedule reflects an overestimate, and some scheduled jobs may have to be terminated (fail) or slowed. If resource capacity increases, the situation is a little better. No jobs need to be disturbed, but the schedule reflects an underestimation, and the scheduler has missed an opportunity to increase goodput. So, answering our question depends on studying schedulers on variable capacity with realistic workloads. We present this question graphically in Figure 3.

Key open research questions include:

1. How do current schedulers respond to capacity variation? (dynamic range, structure, change frequency)
2. How can scheduler performance be improved in these challenging situations?
3. Given the performance, what is a realistic expectation for carbon-emissions benefit from this approach?

3.1 Approach

To characterize the challenge to conventional schedulers, we study workloads and schedulers cloud environments. These workloads are well-known exemplars of their respective environments. For each workload we use a system model that varies the resource capacity available to the scheduler and evaluate performance. Constant resources is a simple model; variable resources can have many different dimensions of variation. We consider three:

- Dynamic range: minimum to maximum capacity
- Variability Structure: random uniform, random walk
- Change Frequency: frequency of capacity variation

Dynamic range captures the distance over which resource capacity varies – from a low to high watermark and back. It is the most foundational element of resource capacity change. Variability structure reflects how capacity is constrained to change from one time period to the next. Such constraints often reflect the realities of physical systems - inductance, momentum, inertia and more – that prevent large instantaneous change. Change frequency reflects our choice to model time discretely – capacity varies only at time period boundaries – so change frequency reflects the size of those periods. In a real system, periods could be defined by external structures (power markets), datacenter physicals (cooling and power sharing control systems), or other factors.

Using these workloads and schedulers, we execute a set of scheduler experiments that explore this multi-dimensional capacity change space, characterizing scheduler performance. In effect, each experiment explores scheduler performance when actual resource capacity diverges from the scheduler’s simple fixed estimate of stable resources. Our goal is both to understand the capabilities of existing state-of-the-art schedulers, and perhaps what constraints or mitigation for capacity variation might be most valuable. With a broad characterization of the negative impacts of capacity variation, we explore several ideas for how to mitigate performance degradation due to capacity variability.

4 Methodology

We define capacity variation, job schedulers, scheduling models, systems, workloads, and metrics used in experiments.

4.1 Dimensions of Capacity Variation

We consider three dimensions of variation, maintaining average resource capacity constant in all cases. These dimensions, dynamic range, structure, and change frequency are illustrated in Figure 4, covering a broad variety of general cases of resource variability. We define them below.

**Dynamic Range** defines distance between maximum and minimum capacity. We consider variation ranges of 0 (constant), 0.2, 0.4, 0.6, 0.8 as a fraction of maximum datacenter capacity. To normalize average capacity at 0.7, this produces dynamic ranges and intervals: 0: [0.7], 0.2: [0.6, 0.8], 0.4: [0.5, 0.9], 0.6: [0.4, 1.0], and 0.8: [0.3, 1.1]. Note that with dynamic range of 0.8, we assume a datacenter has headroom that can temporarily boost over its full 1.0 capacity.
Figure 4. Modeled dimensions of capacity variation include (1) dynamic range, (2) variability structure and (3) change frequency (temporal granularity) on a time-sequence of datacenter capacity from Figure 2.

Structure. defines how much the capacity can change between adjacent time intervals. Random Uniform: Resource capacity in the next interval can be anywhere in the dynamic range and is drawn from a uniform distribution $\mathcal{U}(lbound, ubound)$, appropriate because power prices can be highly volatile. Or, Random Walk: Resource capacity in the next interval can change only by a maximum of stepsize, modeling some continuity from one interval to the next. We use stepsize of one-fourth of the dynamic range throughout the paper.

Change Frequency (Temporal Granularity). defines the frequency of resource capacity changes. Between changes, the capacity is constant. We vary the change frequency from 0.25 per hour (every 240 minutes) to 4 per hour (every 15 minutes).

4.2 Job Scheduling

The resource manager selects a job from the queue and, based on complex priority, selects the resources to run it on. A critical difference in approaches is whether the schedulers consider compute and memory capacity separately, and if oversubscription is practiced.

Oversubscription Scheduling. Among current cloud scheduling models, oversubscription scheduling is widely-adapted to increase system utilization. In this scheduling model, we enforce the CPU and memory limits separately, and over-subscribe the CPU. The idea is to achieve higher resource utilization via statistical multiplexing, exploiting the gap between resource requests, max use, and typical use.

The Borg V2 trace defines the ratio of workload to resources, so we use it unchanged. The Azure trace does not define this ratio so we analyzed the trace and chose an oversubscription of 125% of CPU resources (25% more virtual cores than physical cores), based on the job’s 95th-percentile virtual core utilization instead of requested amount. This choice turns out to match the CPU utilization reported for the Borg V3 traces released April 22, 2020 [58]. In the oversubscription model, jobs can be slowed down when resource capacity shrinks (lower CPU capacity) by the ratio of actual CPU usage to available capacity. Resource capacity reductions decrease the cores available; for the cloud workload, memory capacity is not a limiting factor. Jobs that are slowed down can catch up by claiming surplus CPU capacity in future 5-minute intervals based on actual CPU utilization from the trace.

4.3 Systems

Cloud Systems. We model an Azure cloud cluster that adapts oversubscription scheduling. Therefore, we scale down the hardware to maintain similar system load comparing to non-oversubscription workload. Scaling for both average CPU utilization and the oversubscription rate produces a cloud cluster with 200 nodes (3,200 cores) and 26 TB memory. We also model a Borg cluster with 630 nodes (336 GCU - Google-Compute-Unit) and 300 normalized bytes of memory. This system is sized to match the sampled Borg V2 trace used (Table 2).

4.4 Workloads

We considered a variety of publicly available workloads. The Azure trace was preferred over the Alibaba trace[1] because of its richer batch workload. Azure and Borg traces were chosen as the richest exemplars of cloud or datacenter cluster workloads (key statistics in Table 1).

1This information may not be available in general, but this assumption produces an optimistic estimate.
Cloud workload. We use the Azure and Borg V2 traces. The Azure trace includes VM submission, start, and completion time, requested virtual cores and memory, and actual CPU and memory utilization in 5-minute intervals [19]. As shown in Figure 5 (center), runlength of VMs also ranges from 5 minutes to more than 24 hours (summarized at 24 hours in plot). The largest VMs are only 16 cores. The Borg V2 trace contains task start times, end times, requested CPU and memory, and actual CPU and memory usage in 5-minute intervals. The task runlengths vary from a few seconds to much longer than 24 hours (summarized at 24 hours in plot), parallelism up to 0.5 GCU. Compared to Azure, the Borg trace has more small and short jobs, as well as significant load from long-running jobs (see Figure 5 (right)).

The exact traces used in experiments are summarized in Table 2. The Azure VM traces are run for 7 days, and Borg V2 traces 1-day, sampled. Because they have a significant quantity of jobs, the cloud traces are run for shorter periods than the original trace to limit simulation time. Borg V2 trace has many long-running jobs, so accurate short runs required analysis of typical job length mixes, and careful setup of initial cluster state (a warm start) to get representative results. We validated warm-start with a range of longer (14-day) simulations.

4.5 Metrics
We use two metrics to quantify scheduler performance and Quality of Service (QoS).
- **Goodput** measures the ability of the scheduler to utilize resources to complete jobs. Goodput is the ratio of node-hours for successfully completed jobs to the total available node-hours. Resources that are unscheduled do not count toward goodput.
- **SLO Miss Rate** measures the fraction of jobs that experience SLO violations. In the oversubscription model, Slowdown Rate measures the fraction of jobs that experience a later completion time to the baseline. It quantifies the impact of both oversubscription and resource variability.

5 Impact of Capacity Variation
We evaluate scheduler performance under varying resource capacity to understand how well they can manage variation and when it causes goodput loss for better intelligent power acquisition design. We explore variability dimensions of dynamic range, structure, and change frequency.

5.1 Dynamic Range
First, let’s consider resource capacity variation with different dynamic ranges. In Figure 6, the patterned bar at the left is the scheduler performance (goodput) with same average but no capacity variation – the baseline. The solid blue (left) bars within each group left to right reflect increasing dynamic range, all under random walk structure. Across the clusters, we vary dynamic ranges, from fixed to 0.8, and stepsizes are one-fourth of dynamic range.

With increasing dynamic range, both cloud workloads with oversubscription exhibit little degradation for random walk. We believe this is because the safety margin provided by statistical multiplexing allows much of dynamic capacity change to be absorbed with little negative impact on goodput. Note that the CPU utilization for Borg V2 is much lower than for Azure, as discussed in Section 4.4.

5.2 Variability Structure
We consider two variability structures, random walk and random uniform. In Figure 6, let’s now look at the random uniform case (yellow, right), comparing to no variation (patterned) and random walk results (blue, left), which were also discussed in Section 5.1.

Random uniform also sees little goodput degradation. For Azure workload, in both random walk and random uniform cases, there is a significant goodput increase for large dynamic range case (0.8:[0.3, 1.1]), whereas other dynamic ranges exhibit little goodput difference. It is because 0.8 dynamic range allows large swings in capacity, producing time

---

### Table 1. Key Statistics for Widely Used Public Workload Traces

<table>
<thead>
<tr>
<th>Workload</th>
<th>Azure [19]</th>
<th>Borg V2 [47]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classification</td>
<td>Cloud (VM)</td>
<td>Cluster</td>
</tr>
<tr>
<td>Number of jobs</td>
<td>2,013,767</td>
<td>47,351,173</td>
</tr>
<tr>
<td>Length (days)</td>
<td>30</td>
<td>29</td>
</tr>
<tr>
<td>Job (Avg, StDev) Runtime (hrs)</td>
<td>51.8 &amp; 169</td>
<td>1.84 &amp; 21</td>
</tr>
<tr>
<td>Parallelism</td>
<td>2.6 &amp; 2.4 cores</td>
<td>0.03 &amp; 0.02 NCU²</td>
</tr>
<tr>
<td>Memory</td>
<td>6 &amp; 10 GB</td>
<td>0.03 &amp; 0.02 (Normalized)³</td>
</tr>
<tr>
<td>Year of Trace</td>
<td>2017</td>
<td>2011</td>
</tr>
</tbody>
</table>

### Table 2. Key Trace Statistics for Workload Used in Simulation

<table>
<thead>
<tr>
<th>Workload</th>
<th>Azure'</th>
<th>Borg'</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Jobs</td>
<td>442,784</td>
<td>204,749</td>
</tr>
<tr>
<td>Trace Length (days)</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Job (Avg, StDev) Runtime (hrs)</td>
<td>3.6 &amp; 8.8</td>
<td>0.6, 3</td>
</tr>
<tr>
<td>Parallelism</td>
<td>2.6 &amp; 2.4 cores</td>
<td>0.03, 0.02</td>
</tr>
<tr>
<td>Memory</td>
<td>6 &amp; 11 GB</td>
<td>0.02, 0.02</td>
</tr>
</tbody>
</table>

²The resource unit is rescaled by the largest GCU (Google Compute Unit) capacity of the machines in the traces
³RAM measured in bytes, rescaled by maximum machine memory size in the traces
intervals with higher oversubscription level during capacity decreases with job slowdowns. Moreover, random uniform shows even higher goodput because its structure produces more volatile capacity changes between time intervals. For Borg workload, similar trend but less in quantity due to its lower CPU utilization.

For the cloud scheduling, we conclude that oversubscription scheduling is robust with various variation structure within a moderate dynamic range.

5.3 Change Frequency

Change frequency is another dimension of capacity variation, so we start with a low rate (0.25 changes/hour), and increase to a high rate (4 changes/hour). Note that in all prior experiments used a change frequency of 1 change/hour. We combine change frequency with the other parameters (dynamic range and structure), putting it all together in Figure 7.

Goodputs of both cloud workloads are close to the fixed resource case for all scenarios, and shows no sensitivity to change frequency and dynamic range. Despite of the overall stable performance varying all dimensions, we still see small variations of goodput between change frequency produced by more frequent, disruptive changes of resource capacity. While oversubscription scheduling design can absorb most of these disruptions, intelligent design and close monitoring are still recommended to avoid negative impact. In general, beyond lower parallelism, the flexibility of statistical multiplexing provides greater tolerance of capacity variation.

5.4 Summary

Overall we find that cloud’s oversubscription scheduling model is remarkably robust to capacity change – in almost all dimensions. While with extremely large dynamic range, goodput performance increases significantly, likely eating into the safety margin that allows oversubscription without disrupting SLOs. But overall, cloud system can tolerate wide dynamic ranges, up to 0.8 (80%) of full data center capacity without reduced performance. This is promising indeed for our carbon-emissions aware power acquisition and variable capacity approach.

6 Scheduler Improvements and Effects

We analyze the impact of variable resource capacity on job experiences, and then explore mitigation approaches for the degradation: foresight (advance information) and new scheduling policies.

6.1 Job Slowdowns

We saw in Figure 7 that the cloud oversubscription model exhibits robust goodput. This is in part because this system has a gap between allocation and actual resource consumption. When there are large reductions in capacity or reductions combined with usage spikes, even the oversubscription models can experience a resource deficit that impacts Service-Level-Objectives (SLOs). Figure 6 suggests that large dynamic ranges may eat into the safety margin. When capacity decreases below the safety margin provisioned by the oversubscription resource model, jobs will slow down by the deficit of the resources and latency-sensitive jobs will be severely impacted.

In Figure 8, we show the fraction of jobs experiencing slowdown under resource capacity variation (blue) for Azure workload. While small dynamic ranges do not incur job slowdown, with larger dynamic ranges the capacity decrease cannot be absorbed by the gap. The fraction of jobs that experience slowdown increase drastically with increasing dynamic ranges. Random uniform structure experiences the most slowdowns due to its abrupt changes, with more than half of the jobs are slowed down during extreme dynamic range. On the other hand, random walk structure has a steady growth as dynamic range increases, but still ends up with...
significant job experience impact. So even in the Azure over-
subscription environment, resource capacity variation can
reduce quality of service for latency-sensitive jobs.

6.1.1 Advance Warning (Foresight). With resource vari-
able capacity, some forms of variation may be predictable or
controllable – at a cost. We explore how well schedulers can
explore foresight of resource variability, adapting to an irreg-
ular resource projection. We give the scheduler foresight (a
window of visibility) of zero (baseline) to 24 hours (longest
job duration). Dynamic range is 0.6: [0.4,1.0] and structure
is random walk with stepsize of 0.15. In Figure 9, we look
at goodput results varying length of foresight. For Azure
workload, there is a small benefit, and essentially none for
Borg V2.

Next we consider the impact of foresight on the fraction of
jobs experiencing slowdown (see Figure 10). For all, increased
advance warning reduces job slowdown rates dramatically.
Three hours can eliminate most of the job slowdowns, and
longer foresight does not produce further reductions.

6.2 Intelligent Slowdown Policies
We explore how to best choose the jobs to be slowed down
with the goal of minimizing total slowdowns while main-
taining system goodput. We consider two policies:

- **All Jobs**: Slow down all running jobs by the ratio of
  actual CPU usage to available CPU resource capacity
during which actual CPU usage exceeds available CPU
capacity.
Figure 11. Azure Workload: Fraction of jobs experiencing slowdown versus Intelligent slowdown policy, dynamic range, and structure.

Figure 12. Azure Workload: Goodput versus Intelligent slowdown policy, varying dynamic ranges and structures.

- **Long Jobs**: Slow down only the 10% of running jobs which have the longest times to finish. That is, slow down the job with the longest \((\text{start time} + \text{duration} - t)\) where \(t\) is the current time. Repeat until current CPU usage matches available resource level.

Figure 11 presents results for two slowdown policies (All Jobs, Long Jobs); results show that Long Jobs policy works better, eliminating all job slowdowns (shown as 0%) while dynamic range is small. During extreme cases of large dynamic range, Long Jobs policy is able to reduce job slowdown by up to 50%, drastically improving job experience. This is because jobs that have a long time to run also have a chance to catch up. Figure 12 presents goodput results varying dynamic ranges and variability structures. The results show that while improving job slowdowns, Long Jobs policy can maintain similar or slightly higher goodput performance.

6.3 Summary

We find that resource managers can mitigate the impact of resource capacity variation. For both Azure and Borg V2 here is little goodput degradation, so there is less to fix, but using foresight and slowing longer running jobs can minimize the number of jobs likely to experience an SLO violation, reducing job slowdown rate by 50%.

7 A Case Study on Real Carbon-Driven Load

To illustrate the impact of carbon-based power acquisition and scheduling performance in a real-world scenario, we consider a hypothetical 40-megawatt datacenter operating in the German Power Market[24]. Because the power market varies every day, and has a strong seasonal structure, we pick a set of exemplar days from the 12 most recent months (Sept 2019 - August 2020). When using constant carbon emissions per hour, they have power variation as shown in Figure 13. These twelve days have 24-hour capacity increases from 6% to 16% with an average of 11%.

For each day, we simulate the Azure workload and corresponding cluster and scheduler. This is because the Azure workload shows higher utilization rate and is more sensitive to resource variation as shown in Section 5. We compare a traditional operating mode (fixed power), constant carbon emissions (carbon-emissions-aware), and then add the scheduler enhancements and foresight, graphing goodput in Figure 14. Each cluster of bars depicts the results for a single exemplar day. Shifting from fixed to Carbon-Emission-Aware(CEA) power acquisition produces an increase in goodput as large at 13% on some days and 10.5% on average. The increase of goodput comes from both the increase of total capacity and higher system oversubscription when resource capacity decreases. Figure 15 shows that the increase in goodput comes at the price of more job slowdowns on some days – but for many days the number of slowdowns remains zero.

Finally we consider adding scheduler improvements (see Section 6). Results show that selective job slowdown policy is productive, eliminating the majority of job slowdowns. At the same time, CEA+Selective slowdown outperforms fixed on every one of the 12 days, showing 4% increase of goodput on average. Next we consider further adding advance power.
market information, giving the scheduler 3 hours of foresight into the power levels coming. Results show further but small increment of benefits in increasing goodput and reducing job slowdowns.

Most datacenters have power and cooling headroom\(^4\) and hardware overprovisioning is increasingly popular [50, 51]. However, the quantity is of course limited. To assess these limits on performance and carbon emissions, we show the average goodput for our 12 exemplar days versus capacity headroom (see Figure 16(a)). The goodput remains stable from headroom of 50% and down, whereas the corresponding carbon emissions are significantly reduced by decreasing headroom capacity, by up to 15% (in Figure 16(b)). At all of these levels, the CEA system delivers increased goodput compared to the reference. So it would be fair to scale down quantity purchased to produce a system that delivers the same goodput as fixed capacity. Doing so reduces carbon emission further, delivering a 19% reduction while maintaining system performance.

Finally we assess datacenter power cost, using hourly prices (see Figure 17). The impact of the CEA power acquisition on total power cost is significant, with reductions as large as 14%. This is because low prices are correlated with renewable production. Overall a 20% power cost reduction can be achieved while maintaining the same goodput as of fixed capacity.

8 Related Work

We propose a carbon-emission-aware approach to dynamically acquire power to reduce cloud’s environmental impact; we study resulting performance implications of carbon-driven capacity variations. A great deal of prior research has focused on reducing the carbon footprint of data centers.

\textbf{Optimizing for Green Power Use.} One variant of resource scheduling explores the local management of workload and variable on-site renewable power generation to reduce brown power consumption. These studies consider optimization of criteria such as green-power fraction, workload performance subject to cost, and grid power cost[12, 25, 32] in a system where there is a predictable, local source of renewable power (i.e. solar). [12] dynamically adapts the resource set to the actual workload through shutdown policies to reduce brown energy consumption. Our study assumes externally controlled variable capacity driven by dynamic carbon factor of power market without any presumption of predictability. Our evaluation deals with more general variation which

\(^4\)Headroom is the ability to temporarily exceed these limits safely, and can also exploit thermal inertia.
increases difficulty and also can be extended to capacity changes from other resources.

**Coupling Resource Management with Power Grids.** Zero Carbon Cloud[60] posits the creation of volatile datacenters powered by stranded power (wasted or negative priced renewable power) to produce zero-carbon footprint. The studies view datacenters as wholly on or off. On the other hand, we explore a carbon-driven approach that creates variation in capacity less extreme, and structurally smoother, a much easier and more flexible resource management problem – but still unsolved. Further, we focus on cloud data centers with oversubscription scheduling models. Its robust nature over job usage volatility makes it a good candidate to endure resource variations.

**Large-scale Power Management and Power Capping.** Production data centers have long adapted large-scale power management including power oversubscription and power capping at multiple levels of the power hierarchy. Systems like Facebook’s Dynamo[58] and IBM’s CapMaestro[37] have focused on measuring and budgeting power at server or rack level. Google recently published systems that do power management and shifting at multi-megawatt scale, creating dynamic power constraints for schedulers[49]. These studies do not model schedulers, and interestingly suggest that many scheduling domains give better overall throughput – encouraging variable capacity models from inter-cluster management for scheduling.

Power capping is often framed as known, fixed caps – with variation in application behavior, and managing performance of a set of applications to stay within the caps [22, 36, 42]. In contrast, our framing assumes a dynamic, uncontrolled change in power (resource capacity), due to external factor, the carbon factor of power markets. Potential external factors can be renewable generation, or perhaps the demand in the next datacenter building (unrelated applications or customers perhaps).

**Managing Resource Revocation.** Our resource management approaches dealt with capacity decreases (aka resource reduction or revocation) by intelligently selecting which jobs to terminate. Several systems have done volatile resource management – early work on workstations and PC’s in desktop grids [17, 38] to achieve high throughput on sequential jobs in the face of high rates of individual resource “failure” (revocation), and later work designed to exploit Amazon’s Spot Instances[4] and Google Preemptible VM’s[7]. These systems employed statistical characterization [56, 57] to select appropriate resources, and preventative checkpointing to decrease application “failures” (preserve state across revocations) that have matured into commercial extensions which encapsulate the latency-insensitive, throughput model [3, 33]. These systems are application-oriented; in contrast our work is resource-oriented (cloud).

### 9 Summary and Future Work

We have proposed a carbon-emission-based approach to dynamically acquire power with the goal of reducing carbon emissions in the cloud. This approach reduces both carbon emissions and power cost. The primary challenge is the
variable resource capacity it creates for cloud schedulers. We studied the performance of such schedulers under variation, using commercial cloud workloads. Our results show that oversubscription scheduling model is robust across many dimensions of variation, dynamic range, variability structure, and change frequency. However, dynamic range beyond 0.6 imposes large negative impact on job experiences and should be avoided through design effort.

Exploring mitigation techniques, we find that foresight and slowing longer running jobs minimizes the number of jobs likely to experience an SLO violation. A case study in the German power grid further suggests that proposed dynamic power acquisition approach can significantly reduce carbon emission while maintaining system performance.

These benefits of these techniques suggest that cloud datacenters should adopt CEA power acquisition techniques to reduce both carbon emissions and power costs. Future work includes broader simulations, and studies that couple cloud and power grids as complex systems [35, 49].

Acknowledgments
This work supported by National Science Foundation grants CNS-1405959, CMMI-1832230, and CNS-1901466. Support from Google, Intel, and Samsung is gratefully acknowledged.

References


