ACE: Abstracting, Characterizing and Exploiting Datacenter Power Demands

Di Wang*, Chuangang Ren*, Sriram Govindan†, Anand Sivasubramaniam*,
Bhuvan Urgaonkar*, Aman Kansal‡, Kushagra Vaid†
*The Pennsylvania State University, †Microsoft Corporation, ‡Microsoft Research
{diw5108, cyr5126, anand, bhuvan}@cse.psu.edu, {srgovin, kansal, kvaid}@microsoft.com

***Abstract*—Peak power management of datacenters has tremendous cost implications. While numerous mechanisms have been proposed to cap power consumption, real datacenter power consumption data is scarce. Prior studies have either used a small set of applications and/or servers, or presented data that is at an aggregate scale from which it is difficult to design and evaluate new and existing optimizations. To address this gap, we collect power measurement data at multiple spatial and fine-grained temporal resolutions from several geo-distributed datacenters of Microsoft corporation over 6 months. We conduct aggregate analysis of this data to study its statistical properties. We find evidence of self-similarity in power demands, statistical multiplexing effects, and correlations with the cooling power that caters to the IT equipment.**

With workload characterization a key ingredient for systems design and evaluation, we note the importance of better *abstractions* for capturing power demands, in the form of peaks and valleys. We identify attributes for peaks and valleys, and important correlations across these attributes that can influence the choice and effectiveness of different power capping techniques. We *characterize* these attributes and their correlations, showing the burstiness of small duration peaks, and the importance of not ignoring the rare but more stringent or long peaks. The correlations between peaks and valleys suggest the need for techniques to aggregate and collectively handle them. With the wide scope of *exploitability* of such characteristics for power provisioning and optimizations, we illustrate its benefits with two specific case studies. The first shows how peaks can be differentially handled based on our peak and valley characterization using existing approaches, rather than a one-size-fits-all solution. The second illustrates a simple capacity provisioning strategy for energy storage using the peak and valley characteristics.

***Keywords*-datacenters; power demand characteristics**

I. Introduction

The cost, scalability and environmental concerns arising from the power consumption of datacenters has come under extensive scrutiny. While much of the prior work in the area has looked to reduce energy of computing and cooling systems, the importance of how this energy is dissipated over time (i.e. the power) has gained a lot of recent attention. Power dissipation, particularly the peak or high power draws, impact both operational (op-ex) and capital (cap-ex) expenditures. Electric utilities can charge differentially (op-ex) for peaks (e.g. [4]), especially if such high power draws coincide with high demand across the grid because of supply-demand mismatches that can lead to potential black or brown-outs. Peak power draws also determine the capacity of the power distribution and cooling infrastructure that is provisioned within the datacenter. Prior studies [6], [15] have pointed out that provisioning costs can range between $10-20 per watt, which is incurred even if that watt is not actually consumed.

To address this problem, numerous prior optimizations have been proposed. However, there is a lack of real world datacenter power consumption data to guide the design and enable thorough evaluation of these optimizations. Detailed power consumption data at different temporal scales (from seconds to months) and spatial granularities (from servers, chassis, racks, to datacenters) for datacenters serving important workloads is not easily available. This paper intends to fill this critical void by providing an in-depth analysis of measured power characteristics from the datacenter infrastructure of Microsoft corporation.

Power Characterization: Workload characterization is a key ingredient to the design and analysis of any system that is intended to cater to this workload. It can provide important guidelines regarding how to design the system to handle the average, or a high percentile, of the workload. It also provides the benchmarking ability to evaluate how a given system would perform. More importantly, it can help identify attributes of the workload that stress the system, quantify the statistical properties of these attributes, and exercise these properties in fine tuning and evaluating the system for the future. Such a design can perform much better than one based on just a particular load or trace of the past. The statistical properties enable an easier analysis to quickly examine performance behavior, system design and capacity planning issues, etc., compared to a time consuming design and evaluation loop that may require access to the fine-resolution data over an extensive period of time (that may not necessarily be accessible to everyone).

Recognizing these benefits, there have been several prior efforts at workload characterization for different system design issues, e.g. web, media and cloud services [1], [9], [16], [10], networking [20], [34], file and I/O systems [26], [14], memory system errors [27], impact of datacenter temperature on failures [5], etc., and using these for different optimizations.

While one could take load characteristics and extrapolate them to power demands (using appropriate utilization to power translation models, e.g. [3], [22], [6], [17]), there are

several additional considerations: datacenters host multiple workloads and subsystems, with a complex set of interactions and correlations that could possibly exist across the workloads or subsystems and it is not clear if the extrapolations would hold at the aggregate (temporal and spatial) level. Moreover, power modeling is still an active area of research, with both linear and non-linear correlations between load or utilization and power being suggested [7], [6], and model accuracy may be insufficient for safety critical power capping operations. Instead, direct power measurement based characterization can avoid some of these deficiencies.

Datacenter Power Characterization: Power measurement and characterization, in most prior works, has typically used a few (datacenter) applications and/or a few servers at best. For instance, observations of around 20 servers in a production datacenter in [28] show under-utilization, with the highest power peaks caused by virus scans. A study from IBM [29] examined the temporal and spatial correlation of power consumption in small clusters, each with about 20 servers. A similar characterization of MSN messenger workload [8] has shown opportunities for better provisioning via intelligent workload placement. The most notable large scale undertaking to study datacenter power demands from the provisioning perspective is the published effort from Google [6]. This study identified the headroom for over-provisioning IT equipment within the existing power infrastructure at different spatial scales. The study was more intended to portray the potential of power under-provisioning, rather than as a characterization effort for capturing the statistical properties of the power demands, and their impact on the effectiveness of different power capping and/or power demand shaping knobs. As we will show, a more detailed abstraction (as in our peak and valley attributes) of the characteristics is necessary for these purposes, rather than an aggregate power demand represented as a simple Cumulative Density Function.

To our knowledge, this is the first effort to undertake a systematic characterization of the power consumption of large computing infrastructures, that can be used for the design and evaluation of effective power demand shaping knobs.

Power Capping/Demand Shaping: Broadly, there are three primary categories of power capping knobs which as noted above has both cap-ex and op-ex benefits. Considering the power distribution network as a hierarchy flowing from incoming utility lines, to step-down transformers, UPS units, and Power Distributions Units, that subsequently feed to chassis and racks, and finally to individual outlets for each server, power capping knobs can be employed at one or more of these levels in the hierarchy.Temporal knobs include load scheduling or deferral, which temporally move portions of the load from peaks to valleys to shave the former. These may be implemented through dynamic voltage-frequency scaling (DVFS) techniques that slow down the execution, admission control techniques that drop load during peak demand, or by delaying execution to a low demand time (valleys) [21], [24], [7], [32], [23], [30], [2], [33]. Spatial knobs leverage heterogeneity of power demands at any time across the datacenter, and either statistically multiplexing their low probability simultaneous occurrence to co-locate them within a level [25], [11], [6] or migrate workloads to regions in the hierarchy with headroom (valleys) from regions that are operating at their peak [11], [8], [19], [2]. A recent set of knobs leverages energy storage devices (ESDs), such as batteries, ultra-capacitors, etc., to provide just-in-time extra capacity for power peaks by hoarding required capacity (energy) in previous valleys (when demand was lower) [12], [13], [18], [31]. Depending on where they are placed in the power distribution hierarchy, ESDs can suppress peaks from propagating higher up in the hierarchy.

Overview: In general, the efficacy of all these power capping knobs depends on peak and valley characteristics of power draws in the power hierarchy. For example, temporal deferrals require subsequent valleys large enough to spill over the work from a previous peak. Spatial migration requires a valley elsewhere in the hierarchy to overlap with a peak to be suppressed. ESDs require sufficient valleys, either in number or magnitude, to have sufficient slack to re-charge their capacity preceding the peak that it has to suppress, etc. Consequently, while we do present statistics of aggregate power demands at different temporal and spatial scales (in section II), we focus more detailed characterization results on peaks and valleys in these demands (in section IV), after formally defining these terms for a specified level of power capping in section III.

Contributions: We collected fine-grained power traces from multiple server clusters, each comprising hundreds of servers spanning multiple chassis and racks, across several geo-distributed datacenters of Microsoft corporation over a 6 month duration and make the following contributions:

- We present an aggregate power consumption analysis that shows how power fluctuations change across different spatial and temporal scales. The analysis shows the effect of statistical multiplexing, temporal dependence and self-similarity, and the nature of correlation among server and cooling power consumptions.
- We formally define attributes for peaks and valleys of power demands for a given power cap. These attributes - width (duration), height or depth, and area (energy) - capture the stringency of peaks and slack in valleys.
- We extensively characterize these attributes individually for the peaks and valleys. We also quantify the cross-correlations between them, which would impact the effectiveness of different peak suppression knobs.
- Our analysis shows that there are a large number of peaks of relatively short duration. These short duration

peaks are also typically of small amplitude (height), and occur in bursts. At the same time, we cannot ignore the long duration peaks, which albeit occurring at lower frequencies, impose stringent demands on peak suppression techniques. Our results suggest the possible need to look beyond the immediately successive peak or valley, to perform better aggregate level optimizations, especially for smaller peaks.

- While there are numerous use-cases of such characterization, we explore two illustrative case-studies. The first exploits the properties of peak occurrences and their stringency to differentially employ two peak suppression techniques (load deferral and spatial migration). The second study uses information about peak attributes, and the valley opportunities to come up with rough estimates for ESD technology and capacity provisioning.

II. Aggregate Characteristics

In this section, we present a spatio-temporal analysis of power demand, focusing on its aggregate characteristics.

A. Tracing and Data Collection

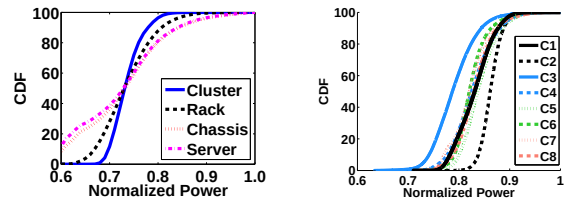
We collected power measurement data from multiple geo-distributed datacenters run by Microsoft corporation over a six month period, between July-December 2011 [1].We specially give results here for data pertaining to 8 representative server clusters (see Table I) in the interest of clarity. Each such cluster comprises several hundreds of servers that span multiple chassis and racks. These clusters run a variety of workloads including web-search, email, Map-Reduce jobs, and several other online cloud applications, catering to millions of users around the world. Each cluster uses homogeneous hardware, though there could be differences across clusters. We name the 8 clusters as C_1, C_2, ..., C_8 and present the trace collection resolution and the type (soft-realtime, batch and interactive) of application that each hosts in Table I. Apart from the IT power of the clusters, we have also collected cooling power.

Cluster Names	Data Resolution	Application Type
C_1, C_2	20 seconds	Soft-realtime, Batch
C_3, C_4, ..., C_8	120 seconds	Interactive

Table I
DATA COLLECTION IS DONE FOR A PERIOD OF SIX MONTHS. EACH CLUSTER HAS SEVERAL HUNDREDS OF SERVERS

B. Statistical Properties

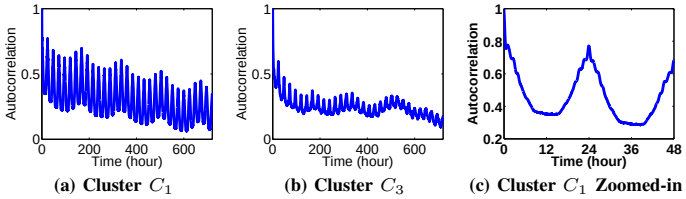
Spatial Characteristics: Figure 1 (a) shows the CDF of power consumption at multiple (spatial) levels for one of the clusters, C_1, starting from cluster-level to rack-level to chassis-level to server-level. The x-axis is normalized with respect to the sum of the maximum demand (over time) of all the servers within that cluster. At the lower levels (e.g. server), there is higher variance in power demands. However, statistical multiplexing effects of these demands, tend to smoothen the fluctuations as we go higher up the hierarchy. For instance, when considering demands at the server level, nearly a third of their duration is spent at a power utilization that is less than 70% of their potential maximum demand. However, the corresponding cluster C_1 rarely drops below its 70% maximum demand. At the other extreme, while the server level power can get over 95% of the maximum demand, the corresponding cluster level power rarely exceeds 80% of the maximum. Thus, there is fairly good statistical multiplexing, and the likelihood of simultaneous peaks across all equipment at the same time reduces as we move higher up in the power hierarchy. These results corroborate observations made in [6], showing the rarity of peak needs at the aggregate level, further motivating the attractiveness for under-provisioning the power hierarchy (especially as we go higher up the hierarchy). Figure 1 (b) shows the CDF of power consumption for the 8 clusters individually. As can be seen, C_1 is representative of a majority of these clusters and hence where intra-cluster details are discussed, we show only this cluster for brevity. The careful reader may observe that C_1's CDF differs slightly in these two graphs, arising because the normalization in (a) is with respect to the potential maximum demand across all its servers, while in (b) it is with respect to its own actual maximum needs over time.



(a) 4 different spatial levels of C_1 (b) 8 clusters C_1 - C_8

Figure 1. CDF of normalized power.

Temporal Characteristics: One way to understand the temporal time series of power demands is through an Auto-correlation Function (ACF) plot with different time-lags. Figure 2 shows the ACFs for clusters C_1 and C_3. While there are some absolute value differences between the two



(a) Cluster C_1 (b) Cluster C_3 (c) Cluster C_1 Zoomed-in

Figure 2. Auto correlation function for different time lags for clusters C_1 and C_3

[1]For business reasons, much of the data is presented as normalized to the relevant maximum, rather than as absolute values.

clusters, the trends are similar, and we specifically focus on C_1, and show a zoomed-in version of its ACF for time-lags stretching up to 48 hours (Figure 2 (c)). While there are significant near-term correlations in the time-series, we note that there is a fairly good time-of-day behavior that is exhibited by the power demands - lags of 24 hours (and multiples) have high correlations, and lags of 12 hours (and its odd number of multiples) are the least correlated. Furthermore, the slower than exponential decay of the ACF indicates that the demands do not follow a Poisson process, with possibility of self-similarity over time. Self-similarity implies structural similarities across a wide range of time scales. To investigate the presence of self-similarity, we calculate the Hurst parameter, using several techniques [9], [20], [34], including variance, R/S method, and periodogram plots. The results are consistent across these techniques. Hurst parameter is a measure of the level of self-similarity with value close to 1 indicating more self-similar. We find a high value for the Hurst parameter, over 0.8, for all clusters (Figure 3), with the log-log variance plot (Figure 5 (a)) and R/S method (Figure 5 (b)) specifically shown for cluster C_1. These quantitatively show the existence of self-similarity in the power demands.

Cluster Name	Hurst Parameter
C_1	**0.93**
C_2	**0.91**
C_3	**0.89**
C_4	**0.90**
C_5	**0.90**
C_6	**0.82**
C_7	**0.87**
C_8	**0.86**

Figure 3. Hurst parameter values of 8 clusters.

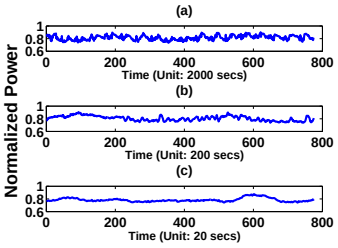
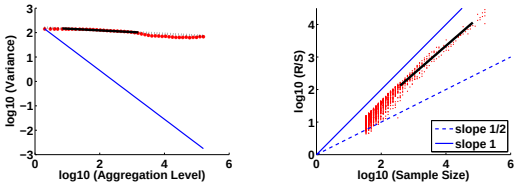


Figure 4. Pictorial view of C_1's power at three different time scales.

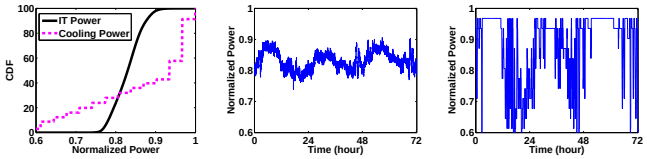


(a) Variance Time Method **(b) R/S Method**

Figure 5. Hurst parameter estimation methods for C_1.

A visual examination of the time series, shown in Figure 4 at different time scales (20s to 2000s) of the average (normalized) power for C_1, also gives some evidence of this behavior. We observe that burstiness persists even at macro time scales, as evidenced by both these results. Consequently, from a datacenter designer's perspective, it is imperative that any methodology employed for peak suppression and power smoothing, recognizes the fact that such power peaks or spikes may occur in close proximity temporally. This impacts the effectiveness of such techniques, e.g., time shifting of load or DVFS may not have enough slack before the next peak, or ESDs would need sufficient time to recharge.

IT and Cooling Power: Figure 6 (a) compares the CDF of IT equipment (servers and networking devices) and cooling power consumptions both normalized with respect to their individual maximum demands for C_1. There are several interesting observations from these results: (i) Cooling power and IT power are correlated as can be seen from Figure 6 (b) and (c). The *Pearson* correlation coefficient between the two is 0.841. However, the cooling power, as is to be expected due to thermal time constants, lags 2 minutes behind the IT power to reach the maximum correlation coefficient of 0.844. As expected based on the high correlation with IT power, the cooling power also exhibits time-of-day behavior and self-similarity. The Hurst parameter value is 0.90. (ii) The variation in cooling power is much more pronounced than that in IT power (also seen visually in Figures 6 (b) and (c)). Beyond its dependence on the IT power draw, cooling power also depends on other parameters including external temperatures, air-flow, etc. High user demand and consequently a high IT power consumption often occurs at times of the day when external temperatures are also among the highest for the day, leading to this wider fluctuation. (iii) The CDF shows that the cooling system is operating closer to its maximum actual draw, much more often than the IT systems. For more than 50% of the time, it is drawing over 90% of its maximum actual draw. This is expected because cooling systems have fewer power states, resulting in more discrete modes of operation.



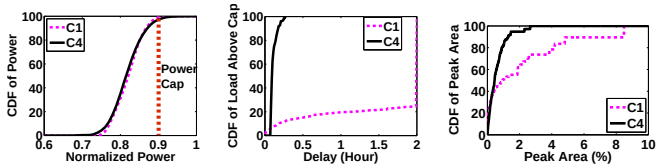
(a) CDF of IT and Cooling Power **(b) IT power** **(c) Cooling power**

Figure 6. Normalized IT and cooling power of C_1.

III. Abstracting Peaks and Valleys

A primary goal of this work is to characterize and analyze power demand time-series into a convenient set of abstractions that facilitate systems design and optimizations for power provisioning, capping, and smoothing (demand shaping). There is a spectrum of abstractions possible, ranging from the very detailed, such as a spatio-temporal reproduction of the entire power demand data at fine resolutions, to a very succinct and possibly simplistic statistic, such as a CDF depicted earlier in Figure 1 (and also used in [6]). As discussed earlier, characterization has wider ramifications than the raw data in many cases, and the raw data may itself

not always be available at the necessary resolutions. But a simple CDF, though attractive, may fall short of the intended purposes when the goal is to design and evaluate power capping and shaping knobs. We illustrate this by taking the power demands of C_1 and C_4, whose power demand CDFs are fairly close as depicted in Figure 7 (a). However, applying a temporal deferring knob such as DVFS to enforce a stipulated power cap (shown as the vertical line in Figure 7 (a)), results in a delay distribution of the deferred load (above the cap) that is very different between these two clusters as in Figure 7 (b). This is because, even if the aggregate area of the power demand above the cap is similar, the distribution of such area across the peaks is very different for these clusters as shown in Figure 7 (c) (peak area will be defined and the consequent performance impact explained shortly).



(a) Power CDF (b) CDF of Load Delay (c) Peak Area CDF

Figure 7. Same power cap imposed on similar CDFs by load deferral, results in different delay distributions.

This reiterates the need for defining, and characterizing, those attributes which would really impact the design and analysis of power capping knobs. Towards this goal, we propose to abstract the power data using power *peaks* and *valleys* that are based on a specified power cap, and identify different attributes for these peaks and valleys. Consider an IT equipment power demand time series p_t, over discretized time $t = 0...T$, in time steps of Δt. This demand could be either from a single server, or a rack, a cluster, or even a datacenter. Let p_{min} and p_{max} be the minimum and maximum power demand over all t in this time series. The dynamic range of power consumption, denoted d, is given by $d = p_{max} - p_{min}$. For any technique employed to optimize the peak power consumption, *the scope of options to set the power cap is only within this dynamic range d. Consequently, we define a power cap c_f, in terms of the fraction, f, (percentage) of this dynamic range. The absolute power cap is then given by* $c_f = (1 - f) \times d + p_{min}$. Our experiments in subsequent sections will explore values of f = 20%, 30% and 40%.

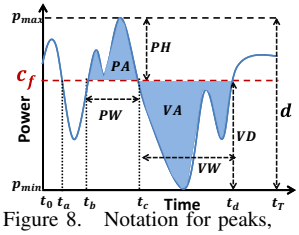


Figure 8. Notation for peaks, valleys, and their attributes.

Setting a cap of c_f to the time series p_t gives rise to peaks and valleys in the power draw as illustrated in Figure 8. For instance, the power draws in the time intervals $[t_0, t_a]$ and $[t_b, t_c]$ are example peaks, while the power draws in time intervals $[t_a, t_b]$ and $[t_c, t_d]$ are example valleys in this figure. Formally, we can define a series $\{i_1...i_k\}$ of points in time where the power demand intersects the horizontal line at a given c_f, i.e. $\forall t \in i_k$, $p_t = c_f$. Note that the power demand in any interval $[i_k, i_{k+1}]$ can be categorized as either a peak or a valley. It is a peak if the power demand within this interval exceeds c_f, and is a valley otherwise. In addition, we also need to consider the extreme cases of intervals $[t_0, i_1]$ and $[i_k, t_T]$ where the beginning and end of time series do not intersect with the horizontal line c_f. These intervals can also be categorized as a peak or a valley depending on whether the power demand in those intervals fall above or below c_f respectively.

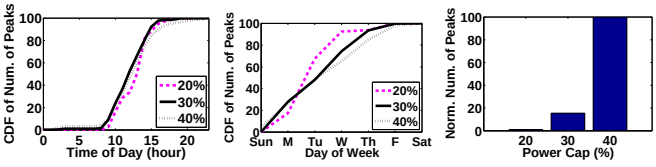
The power demand time series p_t, can now be expressed as a sequence of peaks and valleys of different intervals defined by their respective $[i_k, i_{k+1}]$, with k used to denote the index in the sequence. An interval k, whether a peak or a valley, can be characterized by the following attributes, each of which can have an implication on power capping:

- *Peak Height (PH_k) or Valley Depth (VD_k):* When k is a peak, its height (Power) can be specified as $PH_k = \frac{max_{i_k \le t \le i_{k+1}}\{p_t\} - c_f}{d}$. This is the maximum power draw exceeding the defined cap that needs to be provided over the duration of this peak, normalized as a fraction or percentage of the dynamic power range d. From a practical viewpoint, the magnitude of PH_k would determine the magnitude of the power capping knob that is employed to cap this peak, e.g. number of servers that need to be shutdown, the power states in DVFS to be employed, the capacity of an energy storage device to sustain this peak power need, etc.
  Similarly, when k is a valley, its depth can be specified as $VD_k = \frac{c_f - min_{i_k \le t \le i_{k+1}}\{p_t\}}{d}$. This is the lowest power draw during this valley, capturing its ability to re-charge an energy storage device, take on the load deferred from prior peaks, etc.
- *Peak Width (PW_k) or Valley Width (VW_k):* This is simply the duration (time) of the corresponding peak or valley and is calculated as $i_{k+1} - i_k$. Valley width corresponds to the inter-peak time and vice-versa. These attributes show the frequency of their occurrences, thereby giving an indication of recovery periods between peaks.
- *Peak Area (PA_k) or Valley Area (VA_k):* The area of a peak area PA_k (Energy) is given by $\sum_{t=i_k}^{t=i_{k+1}} (p_t - c_f) \times \Delta t$. Correspondingly, the valley area VA_k is given by $\sum_{t=i_k}^{t=i_{k+1}} (c_f - p_t) \times \Delta t$. The peak area corresponds to the total energy exceeding the power cap, thereby indicating the amount of work that needs to be accommodated by any peak suppression technique without the help of the additional power. The valley area is indicative of how much extra work can be accommodated within the specified cap, e.g. work deferred from a peak, amount of energy for re-charging a storage device, etc.

IV. Characterizing Peaks & Valleys

Having defined the relevant attributes for peaks and valleys, we now characterize their occurrence in our traces. We specially present results for C_1, which is fairly representative. We will mainly focus on f set at 20%, 30% and 40% power caps over the dynamic range, which consequently define the peaks and valleys, in these results.

A. Peak Characterization



(a) Hour of Day (b) Day of Week (c) Peak Reclassification

Figure 9. Peak Occurrences

Figure 9 (a) and (b) show the distribution of peaks for the chosen power caps, for the hours of a day, and days of the week. When we look at the results in the hourly figure, we see a slightly higher distribution of peak occurrences in the working hours, compared to late nights and early mornings as is to be expected. However, with more stringent power caps, there is a *re-classification* of many of the valleys into peaks, i.e. regions which may have been a valley in the 20% cap, may have portions of it re-classified as peaks in the higher caps. This re-classification effect on the relative number of peaks (normalized with respect to the 40% cap) is illustrated in Figure 9 (c). Such re-classification leads to a more uniform distribution of the number of peaks in the 30% and 40% caps, for the day-of-the-week behavior, with the 20% cap being most influenced by day of the week.

Height: Figure 10 shows the CDF of peak heights (PH), across 6 month duration, for the three power caps. Note that a power cap bounds the maximum height of a peak, i.e. the right-most point of the CDF. However, as the figure indicates, a majority of the peaks have small amplitudes. For example, with the 20% cap, nearly 90% of the peaks have amplitudes of 10% or less, which is less than half the cap magnitude. With more stringent caps, while the maximum amplitude can increase, we find that the 30% and 40% caps are still heavily skewed towards the small amplitudes with over 95% of their peaks having amplitudes lower than 10%. This suggests that as amplitudes of peaks already selected by the 20% cap get taller in the 30% and 40% cases, an even larger number of peaks (of smaller amplitudes) are being brought in by the re-classification in these stringent caps, thereby slightly shifting their CDF curves to the left. However, the peaks of amplitudes of 20% or higher are non-zero (though visually not apparent), thereby indicating the need for good peak suppression across a wider range of amplitudes.

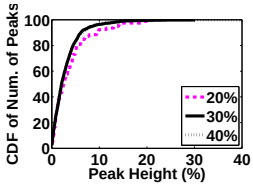


Figure 10. Peak Height Distribution

Width: Figure 11 shows the CDF of the peak width (PW) distribution with the three specified power caps. In the interest of clarity, the x-axis is shown for a zoom-ed in portion of the results. As the caps become more stringent, two factors influence the CDF: (i) more peaks (of smaller widths) get added to the classification, and (ii) existing peaks get wider. When moving from 20% to 30% cap, the former effect is more pronounced, while when moving from 30% to 40%, the rate of addition of new narrow peaks is not sufficient to outweigh the latter widening factor. Regardless of the power caps, these results show that a vast majority of the peaks are quite narrow, i.e. nearly 95% of the peaks last only 4 minutes or less. However, there are a few long peaks as well. For instance, the 20%, 30% and 40% caps have a maximum peak width of 50 minutes, 70 minutes, and 4 hours 30 minutes respectively, identifying the need to handle a wide span of peak durations for any peak suppression technique.

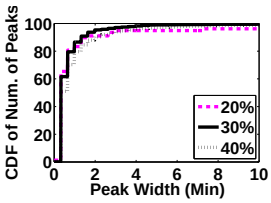
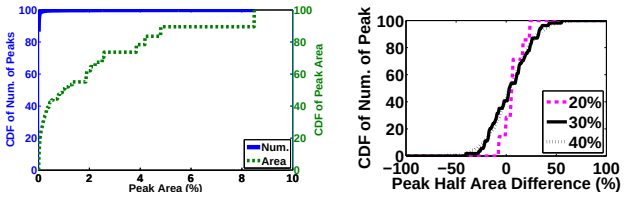


Figure 11. Peak Width Distribution

Area: The area in a peak (PA) can be a more effective measure of the work that needs to be efficiently handled by any peak suppressing strategy, rather than its height or width alone. Figure 12 (a) shows the distribution of this area for the 40% cap (results are similar for the other two). One solid line shows the number of peaks of a certain area (left y-axis), where the peak area (PA) is normalized with respect to the cumulative area of all peaks. As can be seen, this graph accentuates the earlier observations that there are many short and narrow peaks, resulting in a area distribution that is even more skewed towards the left in the CDF. While the longer lasting and taller peaks are very infrequent, as our previous results show, their amplitude and widths can in fact be substantial. This effect is illustrated in the other dashed line of Figure 12 (a) where the right y-axis shows their contribution as a percentage of the total cumulative area under the peaks as a CDF. We notice that the few longer and taller peaks do contribute to over 50% of the total area of peaks - implying that it is extremely important to efficiently handle these rare but demanding peaks.

In addition to the area itself, it may be important to consider the shape of the peaks, i.e. whether it is symmetric or weighed more towards one side (the first half of its duration or the second half). Such information can help understand the impact on performance if workload from these peaks is deferred using temporal knobs such as scheduling or DVFS. For example, if the peak shape is weighted to the left, i.e., it has more work in the earlier phase, then more work will



(a) Peak Area Distribution (b) Peak Symmetry

Figure 12. Peak Area Characteristics

PH_k (%)	PA_k (%)			
	0-0.01	0.01-0.1	0.1-1	≥ 1
0-5	80.30	2.24	0	0
5-10	5.98	8.20	0.25	0
10-20	0.19	1.69	0.76	0.07
≥ 20	0	0.08	0.12	0.12

Table II

DISTRIBUTION OF PEAKS IN TERMS OF THEIR HEIGHT AND AREA (AS % OF PEAKS). AREA IS NORMALIZED WITH RESPECT TO THE TOTAL AREA UNDER PEAKS. $f = 40\%$.

be deferred for a longer duration, than in a symmetric peak. The average deferral delay or loss of performance can hence be greater or lower for an asymmetric peak compared to an equivalent symmetric peak with the same area or amount of work deferred. Figure 12 (b) captures the symmetry of the peaks by plotting the CDF of percentage difference in area between the first and second halves of its duration. Since this difference is more important for longer peaks than short duration ones, we plot this for all peaks lasting longer than 5 minutes. If all such peaks were fully symmetric, there would be a single vertical line at 0%. However, the variances show that there is some amount of asymmetry. However, there are approximately as many peaks with a first half skew as a second half skew, suggesting that on the average, the control effectiveness of peak suppression may even out.

Height vs. Area: Another important characteristic to understand is the height vs. area correlations of peaks, since this has a direct bearing on the choice of technology used in energy storage devices (ESDs) for peak suppression. Certain ESDs such as ultra-capacitors are more efficient for handling a large height (power amplitude), while others such as compressed air energy storage are better for large area (energy). Ragone plots in [31] show significant differences in these efficacies using a 2-dimensional (power vs. energy) plot for different ESD technologies, including ultracapacitors, batteries, and compressed air, etc. Table II shows the percentage of peaks with different height and area ranges for the 40% power cap. While the bulk of peaks are in the small and narrow bin as already observed, we notice the results in this table are weighted more along the diagonal. This suggests that while there could be some vagaries, a large number of peaks are probably suited to a single technology that provides a reasonable trade-off between power vs. energy costs in the Ragone plot (i.e. neither energy biased nor power biased). We will revisit this issue in greater detail in the next section.

B. Valley Characterization

Width: Valley width (VW) measures the inter-peak distance, reflecting the allowable "quiesce" time from the effects of suppressing the previous peak (say using scheduling or DVFS), and is also indicative of the duration for preparing for the next peak (recharging an ESD, migrating workloads, etc.). Figure 13 shows the CDF of the valley width for the 30% power cap (similar for the other caps), and compares it with the corresponding peak width distribution. In the interest of clarity, the x-axis is shown for a zoomed in portion. Valley width is more skewed towards longer durations compared to the peaks. For instance, 60% of peaks last only up to 30 seconds, while 70% of the valleys are of longer duration than 30 seconds. Since the valleys are recovery time or preparation periods between peaks, these results are suggestive of reasonable slack being available for such recovery or prepared-ness. However, the magnitude (area) of the valley needs to be closely examined for more concrete pronouncements as is investigated next. In the interest of space, we merely note that valley depth shows similar distribution as peak height and omit the details for valley depth alone.

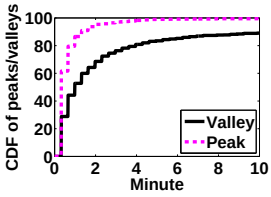
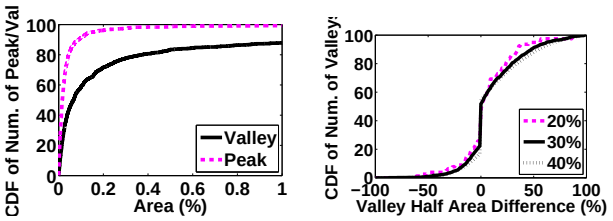


Figure 13. Valley Width Distribution. $f = 30\%$.

Area: The area in a valley (VA) indicates the slack in terms of work (energy) for recovery or prepared-ness between peaks. Similar to peak results, we show the valley area distribution (for a 30% cap) in terms of the number of valleys of a certain area (this is normalized to the total area under peaks so that we can compare peaks and valleys in the same Figure 14 (a)), and the skewness or symmetry of the valley area around its midpoint (in Figure 14 (b)). We also repeat the peak results in these graphs to make direct comparisons. As before, we find that valleys are much more skewed towards larger area compared to peaks. For instance, if we consider the 50-th percentile of peaks, the valley to peak area ratio is roughly 4. When we consider a higher percentile, say 75-th percentile, this same ratio grows to as much as 8. This suggests that on average, a peak is surrounded by a relatively larger (in terms of area) valley suggesting good scope for recovery or prepared-ness. However, one cannot always go by this average case behavior since there are at least a few valleys (0.5% in this case) whose area are smaller than peaks. This motivates the need for a more detailed analysis of peaks and preceding or following valleys, as is discussed in sections IV-C and IV-D.

The valley symmetry figure (Figure 14 (b)) shows a slight positive bias, i.e. more of the valley area occurs earlier in the duration rather than later. This is more preferable, especially for load deferring techniques, which would try to fill up the valley area as early as possible to avoid long delays.



(a) Valley Area Dist. $f = 30\%$. (b) Valley Symmetry

Figure 14. Valley Area Characteristics

VD_k (%)	VA_k (%)			
	0-0.01	0.01-0.1	0.1-1)	≥ 1
0-5	68.71	3.50	0	0
5-10	4.69	14.04	1.47	0
10-20	0.16	1.20	2.68	0.45
≥ 20	0	0.08	0.12	2.91

Table III

DISTRIBUTION OF VALLEYS IN TERMS OF THEIR DEPTH AND AREA (AS % OF VALLEYS). AREA IS NORMALIZED WITH RESPECT TO THE TOTAL OF AREA UNDER PEAKS. $f = 40\%$.

Depth vs. Area: In addition to the valley area and its symmetry, the shape of the valley impacts the ability of an ESD to accumulate enough charge for shaving subsequent peak(s), since technologies limit how fast they can re-charge for a given capacity. To capture such properties, we correlate valley depth with its area (normalized to the total peak area) in Table III, by showing the percentage of total valleys for different depths and area ranges. As peak observations, we find (i) valleys are typically shallow and small (though still larger than peaks as depicted in previous results), and (ii) the table weighted more along the diagonal (again suggesting the possibility of an equal relative importance between power and energy efficiencies of the ESD technologies in the Ragone plot [31] for re-charging).

C. Peak and Following Valley

Having characterized peaks and valleys separately, it is equally important to understand the impact of the characteristics of peak or valley k and the immediately following valley or peak $k+1$. We begin by considering peak k and its next valley $k+1$, where the latter would typically be used by a peak suppression technique such as load deferring and/or DVFS, to utilize the valley area that spills over from the preceding peak. For clarity, we confine our results to the immediately following valley, with cascading carry-overs for several future valleys by these mechanisms possibly having severe performance repercussions. We will, however, discuss such possible occurrences and their implications at the end.

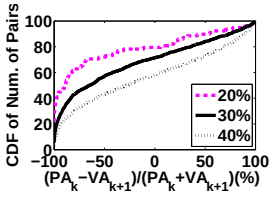


Figure 15. Peak and Following Valley

There are two main metrics for this consideration - (i) the valley width (VW_{k+1}) following the peak width (PW_k) as depicted in Table IV, and (ii) the valley area (VA_{k+1}) following the peak area (PA_k) as captured in Figure 15.

PW_k (min)	VW_{k+1} (min)					
	0-0.5	0.5-1	1-5	5-30	30-60	≥ 60
0-0.5	13.42	13.23	15.45	5.65	1.00	2.28
0.5-1	11.30	8.83	6.64	1.20	0.20	0.48
1-5	9.63	5.65	1.66	0.13	0.01	0.20
5-30	1.88	0.36	0.12	0.01	0.04	0.21
30-60	0.16	0.03	0	0.03	0.01	0
≥ 60	0.16	0	0.01	0	0	0

Table IV

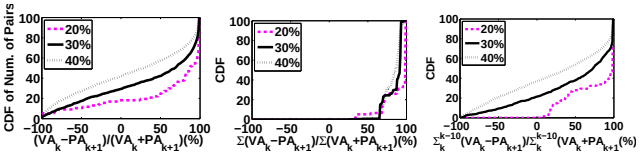
PEAK WIDTH VS. FOLLOWING VALLEY WIDTH (% OF PEAK-VALLEY PAIRS). $f = 40\%$.

Table IV shows the percentage of preceding peak and following valley widths that fall in different duration ranges. It is interesting to view this table or matrix in upper and lower triangular form. The values in the upper triangular part indicate the percentage (roughly 46%) of peak-valley pairs, where the subsequent valley is of longer duration than the peak. The values in the lower triangular part indicate the possibly "worrisome" percentage where the peak is of longer duration than the following valley, i.e. peaks are coming in closer proximation without ample recovery time for load throttling/deferring mechanisms. Around 29% of peak-valley pairs fall in this category.

However, even if the following valley is of shorter duration than the preceding peak, the deferral or DVFS techniques may still perform well if the valley has sufficient area (depth) to accommodate the spilled over load. This is captured in Figure 15 which shows the CDF of the percentage difference between the peak and subsequent valley area. Negative values suggest that the valley area dominate over the peaks, while positive values suggest vice-versa. At 20% and 30% caps, over 70% of the valleys have sufficient room to take on any load deferred from the preceding peak. With a more stringent cap of 40%, as many peaks are larger than their corresponding valleys as vice-versa.

The above two sets of results suggest that any load deferral or DVFS techniques should not presume that the immediately following valley will always have sufficient room to accommodate load spillage from the prior peak. This is particularly true for the short duration peaks as is evident from Table IV. Consequently, such peak suppression mechanisms should recognize the burstiness behavior of these peaks, and employ solutions to address them in a grouped or aggregate manner.

D. Peak and the Preceding Valleys



(a) Immediate Preced. Valley (b) All valleys till that peak (c) Preceding 10 Valleys

Figure 16. Peak and Preceding Valley(s)

The correlations between a peak and its preceding valleys is important for ESD-based peak suppression, which relies

on previous valleys to re-charge its capacity for sourcing power during the current peak. We mainly look at the area (energy) capacity of the valleys for charging opportunities (in the interest of space, we merely note that power capacities are typically adequate).

Figure 16 (a) captures the area difference of a peak and its preceding valley. With 20% caps, nearly 80% of the preceding valleys have sufficient energy charging capacities for the following peak. This percentage decreases to 60% for the 40% cap. However, note that unlike load deferral and DVFS techniques, ESD based peak suppression does not incur performance penalties, and the re-charging does not have to be restricted only to the immediately preceding valley. Figure 16 (b) captures the opportunities for re-charging in all previous valleys, while still discharging for all previous peaks, by showing the CDF of cumulative (from 1 to $k+1$) area difference between each peak and its preceding valley of the entire trace. We note that there are absolutely no negative values in the CDF, indicating that a sufficiently-sized ESD which uses every valley to re-charge, and discharges for every peak, will never run out of capacity to shave any peak in the entire execution. Consequently, these characteristics show a lot of promise for ESD-based peak suppression from a theoretical perspective. However, practically there are several conditions limiting ESD charge and discharge opportunities (some of which will be covered later in section V-B). Hence, we also look at a limited window of opportunity for re-charging, by considering the area differences of 10 consecutive valley-peak pairs in Figure 16 (c). While the 20% power cap imposition can be completely met by the ESDs with this limited window, the results for the 30% and 40% caps are still showing significant negative values. This is again a consequence of our prior observations about the burstiness of relatively small duration peaks which may be separated by even smaller area valleys, leading to fewer opportunities for re-charging to meet the demands. The promise of a much larger window (as in Figure 16 (b)) suggests that an ESD based solution should also look to aggregately suppress these bursty peaks, rather than just-in-time charge and discharge based solutions.

We have also conducted *a time-series analysis of the peaks and valleys alone*, as a ON-OFF process, and not just the aggregate power demand whose results were presented in section II. In the interest of space, detailed results for the peaks and valleys alone are not being presented. We analyzed the tail of the peak and valley distributions, and obtained the corresponding Alpha values (using Hill's estimate) for the ON (peak) process to be 1.2, and the OFF (valley) process to be 1.4 as described in [34]. These values suggest self-similarity in the peak and valley occurrences as well, re-affirming the burstiness and long-tail behavior, and the mandated system design considerations for handling such peaks appropriately as discussed throughout this section.

E. Peaks & Valleys Across Clusters

All of the above analysis focused on a single cluster. Power capping can also benefit from cross-correlation information of peaks and valleys across clusters, to better multiplex the aggregate demand and even migrate the load accordingly. As noted in previous studies [6], [31], power under-provisioning is important at multiple layers of the datacenter power hierarchy, and cluster level capping may become necessary in such cases as opposed to just exploiting multiplexing characteristics at the higher (datacenter) level. We have conducted cross-cluster correlation analysis, and simply summarize the results using two metrics, in the interest of space, for clusters C_1 and C_2, with $f = 40\%$ in both and the numbers are similar across other clusters as well. The first is the probability of a simultaneous peak occurrence in both clusters which we find to be extremely low, i.e. only 0.1% of the time is there a simultaneous peak on both clusters. The second measures whether shifting the peak from one cluster to the other leads to a consequent peak on the latter. We find that this probability is also quite low, with less than 2% of the time that a peak movement to the other cluster results in an exceeding of the cap in the latter. These results suggest the potential of multiplexing which was alluded to some extent in the aggregate characteristics of section II. More importantly, re-distribution of load has tremendous potential in peak suppression, as long as we do not perform such re-distribution too frequently for this to become an overhead by itself. We will illustrate how such spatial differences across clusters, and the temporal analysis of this section, can be used to fine-tune the peak capping knobs in the following section.

V. Exploiting Characteristics

There are several use-cases for exploiting the characteristics quantified in the previous section. We illustrate the utility of such characterization with two case studies. The first exploits the characteristics to fine-tune temporal load deferring and spatial migration knobs, leveraging information about peak-valley behaviors. The second uses the characteristics to come up with a simple capacity provisioning technique for energy storage in the datacenter, to suppress peaks.

A. Tuning Knobs based on Characteristics

As seen in the previous section, both small peaks (which are numerous) and large peaks (though infrequent but extremely demanding) are equally important for peak shaving. One could use a one-size-fits-all policy to shave all the peaks, say using temporal load deferring (LD) to immediate next set of valleys, or spatially moving (SM) them to other clusters. In Table V, we show the impact of LD and SM, if they are uniformly applied to shave all peaks using their individual policies in the columns labeled as LD-only and SM-only for clusters C_1 and C_2 for $f = 40\%$. We capture

their effectiveness in terms of 2 metrics: (i) the 95-th percentile delay of the load above the cap that is being deferred in time, and (ii) percentage of peaks migrated between the two clusters. Note that in an LD-alone scheme, the latter metric will be zero. However, SM-only may still adopt some load deferral if the migrated peak does not find sufficient valley space immediately in the other cluster. While LD-only does not incur any migrations, using it to shave all peaks results in a significant performance penalty. At the other end, blindly migrating all the peaks in SM-only, hardly defers the load, though this would come at a high migration cost (time, bandwidth, locality loss, etc.). The overheads of migration may not be worthwhile for the smaller peaks, as opposed to the larger ones. Hence, one can consider a hybrid scheme LD+SM (see Table V), where LD is used only for small peaks (say up to 5 minute durations). For the peaks, lasting longer than 5 minutes, after applying LD for the first five minutes (since we may not be able to anticipate their durations), the peak is migrated if it still persists. While this does bring down the migrations substantially, the load deferral delays are still quite significant (albeit smaller than LD-only). Finally, we can leverage our observations from the previous section to explain and further optimize this hybrid approach. Recall that despite the majority of peaks being short, a majority of them also had a subsequent valley which could not accommodate all of the deferred load. Such burstiness led us to believe that an aggregate level optimization may be more productive. Consequently, we fine-tune the hybrid approach to take several of these small peaks and migrate them in an aggregated manner (AGG-SM) while migrating the larger peaks if they persist beyond 5 minutes. Results for this approach are shown in the column labeled LD+AGG-SM of Table V. By such fine-tuning of the hybrid approach, the migrations have dropped substantially, and the load deferral delays are quite comparable to the SM-only approach, thus performing better than LD-only and SM-only approaches individually.

	PW (min)	LD-only	SM-only	LD+SM	LD+AGG-SM
Peak Type	PW=0-1	LD	SM	LD	LD/AGG-SM
	PW=1-5	LD	SM	LD	LD/AGG-SM
	PW=5-30	LD	SM	LD/SM	LD/SM
	PW≥30	LD	SM	LD/SM	LD/SM
Result	95th percent. delay (hour)	12	0.4	2	0.4
	% of peaks migrated	0	100	3	20

Table V
TUNING KNOBS BASED ON CHARACTERISTICS

B. ESD Provisioning

ESDs have come under recent scrutiny as a peak suppression mechanism, to provide just-in-time power when caps are being exceeded. However, there is a diversity in ESD technologies that is captured by a Ragone plot, with different technologies suited or cost-effective for different kinds of peaks. For instance, a prior study [31] has shown ultra-capacitors to be attractive for power-intensive tall and narrow peaks, and compressed air for energy-intensive broad peaks. Batteries fall in the middle of this spectrum. However, as the prior study has shown [31], there are numerous considerations when provisioning these ESDs for peak suppression: power needs, energy needs, ability to charge and discharge in a given time, lifetime degradation due to repeated charge and discharge cycles, ramp rate, energy losses, etc. Taking all these factors, and provisioning the ideal capacity for a given datacenter's power profile over an extensive period of time (such as the 6 month, 20 second resolution), is a cumbersome exercise. Further, considering a single time-series for such provisioning may not necessarily capture all the statistical properties of the datacenter's behavior, which is in fact more important for future demands (rather than just looking at the past history). Our characterization results, presented in the previous section, can provide one possibly simple methodology for ESD provisioning, even if it is not the optimal, as discussed below.

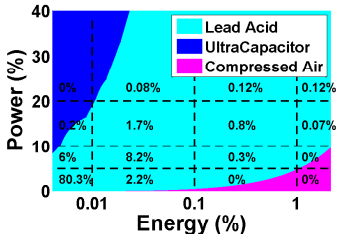


Figure 17. Most cost-effective ESDs for Energy-Power Needs. Energy is expressed as percentage of the area under the peaks, and power is expressed as PH. On this graph, we overlaid information from Table II to illustrate the suitability of each ESD for the different kinds of peaks.

Choosing Technology: Given a peak of height PH and area PA, the capacity of an ESD and hence its cost (C) to shave this peak can be determined by its technology's power (C^P) and energy (C^E) density cost as $C = max\left(PH \times C^P, PA \times C^E\right)$. Using this simple model, we compute the (PH, PA) regions over which 3 ESD technologies (Ultra-capacitor, Lead acid battery, and Compressed air energy storage) under consideration are the most cost-effective as in Figure 17. We can overlay the PH and PA distributions from Table II on this figure, to examine the suitability of each technology to shave a certain number of peaks. We see that a majority (99.8%) of peaks from Table II falls in the region where lead-acid battery is the most cost effective technology. Very few (0.2%) of the peaks fall in the ultra-capacitor region, and this constitutes only about 0.1% of the area of the peaks. Consequently, we can simply go with the lead-acid battery option, amongst the 3 technology choices based on our characteristics results.

Quantifying Capacity: The shape - height which indicates peak power, and area which indicates energy needs - of the peaks determines the required ESD capacity for any given technology. We, thus, examine the characteristics in

Figure 10 and Figure 12 (a). Rather than go for the 100-th percentile, we pick the knees of these curves - 90th percentile of the peaks in terms of height, peaks contributing to the 90% of total area under peaks - with the former indicating the power capacity and the latter representing the required energy capacity. The maximum between these two is what needs to be provisioned. However, we also need to ensure that there is sufficient slack in the valleys to re-charge for this capacity. For this purpose, we examine Figure 16 to examine the re-charging opportunities. While the immediately preceding valley(s) may not have enough slack for such re-charge (Figures 16 (a) and (c)), the results in Figure 16 (b) suggests that greedily re-charging at every opportunity may provide the slack to re-charge for this capacity to suppress all peaks.

Using these rules-of-thumb as a heuristic for ESD provisioning, in Table VI we show the capacity selected by this heuristic for 2 specific ESD technologies - lead-acid batteries, and ultra-capacitors, showing its effectiveness in shaving the peaks (both number and area). We compare these results with an Optimal capacity provisioning algorithm, that is guaranteed to provide the minimal capacity to shave all peaks in the given data. While the latter does need to extensively run through the time series of power demands, to ensure these guarantees (minimal capacity, charge/discharge rate guarantees, account for energy losses, etc.), we find that our simple heuristic approach shaves over 99% of the peaks, and nearly 90% of the peak area, with a capacity (and corresponding cost), that is less than half of the capacity (for lead-acid batteries) than what the Optimal algorithm specifies. As is evident in Figure 17, lead-acid is an overwhelming favorite as far as technology is concerned.

Tech.	Approach	Capacity (% peak area)	Cost (% of Opt LA)	Peaks shaved (% of total peaks)	Area shaved (% of total peak area)
LA	Heuristic	6.0	47	99.97	89.42
LA	Opt.	12.1	100	100	100
UC	Heuristic	6.0	789	99.97	89.43
UC	Opt.	10.8	1494	100	100

Table VI
ESD CAPACITY PROVISIONING FOR LEAD ACID (LA) BATTERY AND ULTRA-CAPACITOR (UC).

VI. CONCLUDING REMARKS & FUTURE WORK

We have undertaken a detailed characterization of power measurements of geo-distributed datacenters of Microsoft corporation at fine temporal and spatial resolutions over a 6 month duration. Aggregate analysis of such raw data shows (i) statistical multiplexing of power demands that can enable more aggressive under-provisioning at higher layers of the power hierarchy; (ii) significant evidence of self-similarity in the power demands, together with time-of-day behavior; and (iii) correlations between the IT and cooling power, with the latter showing higher variance, and a 2-minute lag behind the former. While these aggregate characteristics can be useful by themselves, there is a need for better abstractions to design, evaluate, and fine-tune peak suppression mechanisms to achieve a desired level of power capping. Towards this goal, this paper has made the following contributions:

Abstractions: We have formally defined peaks and valleys, their important attributes (height, width, area), and the correlations between peaks and valleys that need to be studied towards designing and understanding the potential of any peak suppression mechanism.

Characterizing Peaks and Valleys: We have extensively characterized peak and valley attributes individually, and their cross-correlations. Results show that while there are an overwhelming number of small duration and small amplitude peaks, we cannot afford to ignore the few large ones that have very stringent demands. While on average, valleys do offer enough slack for load deferment or peak preparation, there are bursts of peaks which do not have sufficient valleys immediately following or preceding them. Further, there is significant potential of migrating load to exploit peaks and valleys across clusters, as long as we can restrict the number of such migrations to avoid the consequent performance penalties. These suggest aggregated optimizations of peaks and valleys.

Exploiting Characteristics: There are numerous use-cases for our characteristics, and we illustrated two specific case-studies in the limited space of this paper. The first used the characteristics to fine-tune load deferring and migration based on the kinds of peaks, in an aggregated manner. The second showed a simple approach to energy storage provisioning that only uses aggregate characteristics, rather than an extensive approach considering every possible eventuality in the entire power demand time series.

There are several more opportunities for future work to leverage the proposed characterizations, including further use of predictability of peak and valley characteristics for fine-tuning peak suppression knobs, synthesizing workloads with the broad statistical properties that we have identified, analytical models to work with the characteristics for quick performance and capacity provisioning estimates, energy supply side sourcing and management issues (including renewables and cost) as opposed to just demand-side capping.

ACKNOWLEDGMENTS

This work was supported by NSF grants 0811670, 1152479, 1205618, 1213052, 1147388, 1302225, 1302557 and CAREER award 0953541.

REFERENCES

[1] M. F. Arlitt and C. L. Williamson. Web server workload characterization: the search for invariants. In *Proceedings of SIGMETRICS*, 1996.

[2] R. Bianchini and R. Rajamony. Power and Energy Management for Server Systems. *Computer*, 37(11), 2004.

[3] D. Brooks, V. Tiwari, and M. Martonosi. Wattch: a framework for architectural-level power analysis and optimizations. In *Proceedings of ISCA*, 2000.

[4] Duke utility bill tariff. http://www.duke-energy.com/pdfs/scscheduleopt.pdf.

[5] N. El-Sayed, I. A. Stefanovici, G. Amvrosiadis, A. A. Hwang, and B. Schroeder. Temperature management in data centers: why some (might) like it hot. In *Proceedings of SIGMETRICS*, 2012.

[6] X. Fan, W.-D. Weber, and L. A. Barroso. Power Provisioning for a Warehouse-sized Computer. In *Proceedings of ISCA*, 2007.

[7] A. Gandhi, M. Harchol-Balter, R. Das, and C. Lefurgy. Optimal power allocation in server farms. In *Proceedings of SIGMETRICS*, 2009.

[8] L. Ganesh, J. Liu, S. Nath, G. Reeves, and F. Zhao. Unleash Stranded Power in Data Centers with RackPacker. In *Workshop on WEED*, 2009.

[9] M. W. Garrett and W. Willinger. Analysis, modeling and generation of self-similar VBR video traffic. In *Proceedings of SIGCOMM*, 1994.

[10] D. Gmach, J. Rolia, L. Cherkasova, and A. Kemper. Workload analysis and demand prediction of enterprise data center applications. In *Proceedings of IISWC*, 2007.

[11] S. Govindan, J. Choi, B. Urgaonkar, A. Sivasubramaniam, and A.Baldini. Statistical profiling-based techniques for effective power provisioning in data centers. In *Proceedings of EUROSYS*, 2009.

[12] S. Govindan, A. Sivasubramaniam, and B. Urgaonkar. Benefits and Limitations of Tapping into Stored Energy For Datacenters. In *Proceedings of ISCA*, 2011.

[13] S. Govindan, D. Wang, A. Sivasubramaniam, and B. Urgaonkar. Leveraging Stored Energy for Handling Power Emergencies in Aggressively Provisioned Datacenters. In *Proceedings of ASPLOS*, 2012.

[14] A. Gulati, C. Kumar, and I. Ahmad. Modeling workloads and devices for IO load balancing in virtualized environments. *SIGMETRICS Perform. Eval. Rev.*, 37(3), 2010.

[15] J. Hamilton. Internet-scale Service Infrastructure Efficiency, ISCA Keynote 2009.

[16] A. K. Iyengar, M. S. Squillante, and L. Zhang. Analysis and characterization of large-scale web server access patterns and performance. *World Wide Web*, 2(1-2), 1999.

[17] A. Kansal, F. Zhao, J. Liu, N. Kothari, and A. A. Bhattacharya. Virtual machine power metering and provisioning. In *Proceedings of SOCC*, 2010.

[18] V. Kontorinis, L. E. Zhang, B. Aksanli, J. Sampson, H. Homayoun, E. Pettis, D. M. Tullsen, and T. S. Rosing. Managing Distributed UPS Energy for Effective Power Capping in Data Centers. In *Proceedings of ISCA*, 2012.

[19] K. Le, R. Bianchini, M. Martonosi, and T. Nguyen. Cost- and energy-aware load distribution across data centers. In *Workshop on HotPower*, 2009.

[20] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson. On the self-similar nature of ethernet traffic (extended version). *IEEE/ACM Trans. Netw.*, 2(1), 1994.

[21] K. Ma, X. Li, M. Chen, and X. Wang. Scalable power control for many-core architectures running multi-threaded applications. In *Proceedings of ISCA*, 2011.

[22] J. C. McCullough, Y. Agarwal, J. Chandrashekar, S. Kuppuswamy, A. C. Snoeren, and R. K. Gupta. Evaluating the effectiveness of model-based power characterization. In *Proceedings of USENIX*, 2011.

[23] D. Meisner, C. M. Sadler, L. A. Barroso, W. Weber, and T. F. Wenisch. Power management of online data-intensive services. In *Proceedings of ISCA*, 2011.

[24] R. Raghavendra, P. Ranganathan, V. Talwar, Z. Wang, and X. Zhu. No Power Struggles: Coordinated multi-level power management for the data center. In *Proceedings of ASPLOS*, 2008.

[25] P. Ranganathan, P. Leech, D. Irwin, and J. Chase. Ensemble-level Power Management for Dense Blade Servers. In *Proceedings of ISCA*, 2006.

[26] M. Satyanarayanan. A study of file sizes and functional lifetimes. In *Proceedings of SOSP*, 1981.

[27] B. Schroeder, E. Pinheiro, and W.-D. Weber. DRAM errors in the wild: a large-scale field study. In *Proceedings of SIGMETRICS*, 2009.

[28] A. Vasan, A. Sivasubramaniam, V. Shimpi, T. Sivabalan, and R. Subbiah. Worth their watts? - an empirical study of datacenter servers. In *Proceedings of HPCA*, 2010.

[29] A. Verma, G. Dasgupta, T. K. Nayak, P. De, and R. Kothari. Server workload analysis for power minimization using consolidation. In *Proceedings of USENIX*, 2009.

[30] A. Verma, P. De, V. Mann, T. Nayak, A. Purohit, G. Dasgupta, and R. Kothari. Brownmap: Enforcing power budget in shared data centers. In *Proceedings of MIDDLEWARE*, 2010.

[31] D. Wang, C. Ren, A. Sivasubramaniam, B. Urgaonkar, and H. Fathy. Energy storage in datacenters: what, where, and how much? In *Proceedings of SIGMETRICS*, 2012.

[32] A. Weisel and F. Bellosa. Process cruise control-event-driven clock scaling for dynamic power management. In *Proceedings of CASES*, 2002.

[33] A. Wierman, L. L. H. Andrew, and A. Tang. Power-aware speed scaling in processor sharing systems: Optimality and robustness. *Perform. Eval.*, 69(12):601–622, 2012.

[34] W. Willinger, M. S. Taqqu, R. Sherman, and D. V. Wilson. Self-similarity through high-variability: statistical analysis of ethernet lan traffic at the source level. In *Proceedings of SIGCOMM*, 1995.