Title
Battery Provisioning and Associated Costs for Data Center Power Capping

Authors
Kontorinis, Vasileios
Sampson, Jack
Zhang, Liuyi Eric
et al.

Publication Date
2012-07-30

Peer reviewed

UCSD Technical Report, August, 2012

Battery Provisioning and Associated Costs for Data Center Power Capping

Vasileios Kontorinis Jack Sampson Liuyi Eric Zhang Baris Aksanli
Houman Homayoun Tajana Simunic Rosing Dean M. Tullsen

Department of Computer Science and Engineering, UC San Diego

Abstract

Power over-subscription can reduce costs for modern data centers. However, designing the power infrastructure for a lower operating power point than the aggregated peak power of all servers requires dynamic techniques to avoid high peak power costs and, even worse, tripping circuit breakers. This work presents an architecture for distributed per-server UPSs that stores energy during low activity periods and uses this energy during power spikes. This work advocates explicit sizing of the distributed UPS batteries for power capping and provides a methodology to properly select the properties of the battery in order to reduce total datacenter cost of ownership per server. Depending on workload diurnal patterns and oversubscription assumption this approach can achieve up 5-10% reduction in overall costs per server.

1 Introduction

The costs of building and running a data center, and the capacity to which we can populate it, are driven in large part by the peak power available to that data center. This work demonstrates techniques to reduce the observed peak power demand for data centers with distributed UPS batteries, enabling significant increases in data center capacity and reductions in cost.

Modern data center investments consist of one-time infrastructure costs that are amortized over the lifetime of the data center (capital expenses, or capex) and monthly recurring operating expenses (opex) [19]. Capex costs are proportional to the provisioned IT power per facility, estimated at $10-20 per Watt [6, 30, 39], as each Watt of computing power requires associated support equipment (cooling, backup, monitoring, etc.). Utilities typically charge a power premium that is tied to the peak power. This can become a significant portion of the monthly bill, up to 40% [15]. This paper examines the use of distributed batteries in the data center to reduce both capex and opex costs.

Power infrastructure is commonly over-provisioned in data centers to accommodate peaks and to allow for future expansion. However, to improve common case utilization, we can intentionally over-subscribe (under-provision) the power infrastructure [10, 19, 21, 23, 28]. Over-subscribing provisions power infrastructure to support a lower demand than the largest potential peak and employs techniques to prevent power budget violations. In the worst case, such violations could trip circuit-breakers and disable whole sections of the data center, causing costly down time. To avoid this, data centers can employ power capping approaches such as CPU capping, virtual CPU management, and dynamic voltage and frequency scaling (DVFS) [23, 25, 31]. CPU capping limits the time an application is scheduled on the CPU. Virtual CPU management limits virtual machine power by changing the number of virtual CPUs. DVFS attacks the peak power problem by reducing chip voltage and frequency. However, all of these techniques result in performance degradation. This is a problem for any workload that has performance constraints or service-level agreements because power management policies apply these performance-reducing mechanisms at the exact time that performance is critical – at peak load.

Govindan, et al. [15] introduce a new approach that has no performance overhead in the common case. They leverage the energy stored in a centralized data center UPS to provide energy during peak demand, effectively hiding the extra power from the power grid. This technique is shown to work well with brief (1-2 hours), high-magnitude power spikes that can be completely "shaved" with the energy stored in batteries; however, it is less effective for long (8-10 hour) spikes. For longer spikes, they suggest a hybrid battery-DVFS approach.

However, many large data centers do not employ centralized batteries. Distributed, per-server batteries represent a more economical solution for battery backup. They scale naturally with the data center size and eliminate a potential single point of failure. Google employs this topology in their state-of-the-art data centers [13].

When leveraging a distributed UPS architecture to shave peak power, challenges arise due to the lack of heavy over-provisioning and the distributed nature of the batteries. The absence of over-provisioned UPSs means we need to justify the use of larger batteries based purely on cost savings from power capping. We need policies to determine how many batteries to enable, which batteries to enable, and when. However, there are also opportunities compared to prior solutions. In a centralized UPS architecture, all power typically comes from either the battery or the utility. Thus, when batteries are enabled, they supply all datacenter power and drain quickly – if we only supply the over-threshold power, the batteries can sustain longer peaks. This is easily done in the distributed architecture by simply enabling enough batteries to hide the desired peak.

In this work, we discuss the applicability of battery-enabled power capping to distributed UPS topologies. We make the following contributions:

1. We provide a thorough total cost of ownership analysis for modern data centers with focus on the power infrastructure costs. Using this model we quantify the benefits of power oversubscription in terms of total cost of ownership per server. We discuss how these benefits change with oversubscription applied at different levels of the power hierarchy, and we study power oversubscription for a range of costs and power bugets per server.

2. We provide a methodology to size the distributed UPS batteries, compare alternative battery technologies and demonstrate how to optimally select user-defined battery parameters such as the depth of discharge.

3. We argue that investing money in distributed UPSs by explicitly provisioning UPS batteries for peak power capping enables higher degree of power oversubscription and reduces overall costs per server.

This paper is organized as follows. Section 2 presents common UPS topologies and the associated trade-offs. Section 3 describes our total cost of ownership analysis and section 4 quantifies the benefits of power oversubscription. In Section 5 we contrast alternative battery technologies for frequent battery charge/discharge in the data center context and elaborate on their properties. Section 6 reviews related work in power capping techniques and Section 7 concludes.

2 Background

Primary power delivery in data centers is through a utility line. Data centers are also equipped with a diesel generator unit which acts as a secondary source of power during a utility failure. To facilitate switching power between the utility and the diesel generator, an automatic transfer switch (ATS) selects the source of power, which takes 10-20 seconds [15]. During this short and critical interval, the UPS units supply the power to the data center. In the centralized topology shown in figure 1(a), the power from a single UPS is fed to several Power Distribution Units (PDUs) to route the power to racks and servers. To eliminate the transfer time of the power line to the UPS, data centers commonly use double conversion UPSs. With double conversion UPSs, power is transformed from AC-to-DC to be stored in batteries and then from DC-to-AC to be used by the racks and servers. Although this organization has zero transfer time to the UPS (the UPS is always in the power path), the availability of the whole data center is dependent on the UPS. Additionally, double conversion introduces 4-10% power losses during normal operation [13].

The centralized UPS topology in figure 1(a) does not scale well for large data centers. This topology either requires double conversion, so that the power network distributes AC power, to be converted again to DC, or it distributes DC over the large network, resulting in higher cable losses. The inefficiency of AC-DC-AC conversions becomes more costly at scale. The UPS is also a single point of failure and must be overprovisioned.

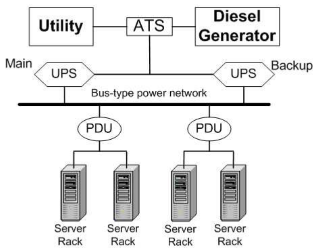
Figure 1(b) shows the distributed design adopted by Facebook. A cabinet of batteries for every 6 racks, or a total of 180 servers, replaces the centralized UPS [9]. This design avoids double conversion by customizing the server power supply unit to support both AC power (from the grid) and DC power (from the battery cabinet). DC power is distributed from the UPS to the servers, but in this case that is a much shorter distance. Google goes even further, attaching a battery on every server after the Power Supply Unit (PSU) [13], as depicted in figure 1(c). This design also avoids the AC-DC-AC double conversion, saving energy under normal operation, and brings the AC distribution even closer to the IT load, before it is converted.

Availability in data centers is a function of how often failures happen, the size of the failure domain, and the recovery time after each failure. UPS placement topology impacts the availability of the data center, particularly the associated failure domain. The more distributed the UPS solution, the smaller the failure domain. Thus, the centralized design requires full redundancy, while the Google approach provides none (loss of a single node is insignificant), further reducing cost.

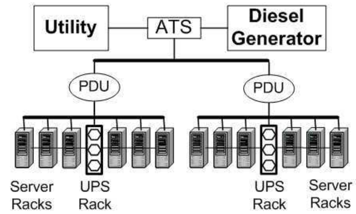
3 Total Cost of Ownership analysis

Modern data centers are typically power limited [39]. This means that the overall capacity (number of servers) is limited by the initial provisioning of the power supporting equipment, such as utility substations, diesel generators, PDUs, and cooling. If we reduce the peak computing power, we can add additional servers while remaining within the same power budget, effectively amortizing the initial investment costs over a larger number of servers. Moreover, extra work done per data center should result in fewer data centers, greatly reducing capex costs.

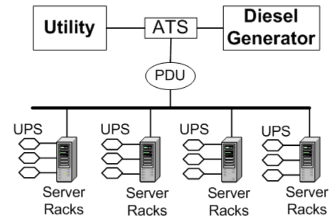
Distributed UPSs are currently designed to support the whole computing load long enough to ensure safe transition from the main grid to the diesel generator. This time window (less than one minute) translates to batteries with insufficient stored energy for meaningful peak power shaving. Therefore, to enable peak power capping using UPS stored energy in the distributed context, we need to overprovision per server battery capacity. Since this requires a greater investment compared to current systems, we must establish that doing so reduces total costs more than the cost



(a) Centralized



(b) Distributed (Facebook)



(c) Distributed (Google)

Figure 1: Power hierarchy topologies with (a) centralized UPS and (b,c) distributed UPS solutions.

$$\begin{aligned} TCO/server =& (dataCenterDepreciation + dataCenterOpex + serverDepreciation + serverOpex)/N_{servers} \\ =& ((FacilitySpaceDepr + UPSDepr + PowerInfrastructureDepr + CoolingDepr + RestDepr) + dataCenterOpex \\ &+ serverDepr + (ServerRepairOpex + (ServerEnergyOpex + ServerPowerOpex) * PUE))/N_{servers} \end{aligned} \tag{1}$$

of providing additional battery capacity. This section discusses the TCO model we use to examine the battery overprovisioning that makes financial sense and maximizes total profits. We first quantify the benefits of ideal power oversubscription – placing more servers without accounting for extra cost in power management techniques – and then we incorporate the investment in required battery capacity to avoid power budget violations.

The profitability of an investment is defined as the generated revenue minus the associated total cost of ownership (TCO). The data center revenue equals the number of servers times the income per server. We assume constant income per server. Therefore, maximizing the profitability per server is equivalent to minimizing the TCO per server. Our TCO analysis is inspired by the data center cost chapter in Barroso and Holzle [19]. For simplicity, we assume sufficient initial capital, hence there are no monthly loan payments, and full capacity for the data center (limited by the provisioned power) from the first day. Financing the investment should similarly scale all TCO components and should not affect our results. Reducing the deployment time of the data center is a problem orthogonal to reducing capex costs and will not be treated here. The TCO/server is given by equation 1.

In this equation, data center depreciation is the monthly depreciated cost of building a data center (we assume 10 year straight-line depreciation [19]) The assets required for a data center are land, UPS and power infrastructure (diesel generators, PDUs, back-room switchgear, electrical substation), cooling infrastructure (CRAC, economizers), as well as several other components such as engineering, installation labor, racks, and system monitors that we include in *RestDepreciation*. The data center opex is the monthly cost for running the data center (infrastructure service, lighting). We collect the depreciation and opex cost information for a data center with 10MW provisioned computing power (critical power) from APC's commercial TCO calculator [3].

Servers typically have shorter lifetimes and are depreciated over 4 years. Server opex consists of server repairs and the electricity bill. Utility charges have a power component and an energy component. The power component is based on the peak sustained power for a 15 minute window over the period of a month [5] while the energy is based on the total data center energy used (different charging models provide similar results). To account for the electricity consumed by infrastructure, excluding servers, we scale the total server peak power and energy by the power usage effectiveness (PUE), assumed at 1.15 [13]. We assume a customized commodity server similar to Sun Fire X4270, with 8 cores (Intel Xeon 5570) at 2.40 GHz, 8 GB of memory, and costing $1500. The inputs to our TCO model are summarized in table 1.

Data center Critical Power	10 MW
Server	Idle Power: 175W, Peak Power: 350W (measured)
Number of servers	28000 (critical power / server peak)
Average Server Utilization	50% [19]
Utility Prices	Energy: 4.7 c/KWh, Power: 12 $/KW [5, 15]
Server cost	$1500
PUE	1.15 [13]
Amortization Time	Infrastructure: 10 years, Servers: 4 years [19]

Table 1: TCO model assumptions

The table and the pie chart in Figure 2 show the breakdown of TCO/month/server. The major TCO component is server depreciation (40.6%). Infrastructure related components (facility space, power, cooling, and data center opex) account for more than 35%. In the same table, we also present how the ratio of each TCO component per server changes when we are able to add additional servers within the same power budget. Server depreciation, server opex, and UPS TCO scale with the number of servers and are constant. The energy component of the utility bill also scales with the number of servers, but the power component stays the same and is amortized over more servers. Infrastructure costs are also amortized over a larger number of servers. The UPS cost (estimated as the total cost of the server-attached batteries) represents a very small portion of the TCO; it is marginally visible in the pie chart. Our pro-

TCO component	TCO/month (TCO/month/server)	TCO/server trend with extra servers
Facility Space depreciation	96,875$ (3.46$)	Decreasing
UPS depreciation	3,733$ (0.13$)	Constant
Power Infrastructure depreciation	169,250$ (6.04$)	Decreasing
Cooling infrastructure depreciation	70,000$ (2.50$)	Decreasing
Rest depreciation (racks,monitoring,engineering,installation)	255,594$ (9.13$)	Decreasing
Data center opex (maintenance, lighting)	213,514$ (7.63$)	Decreasing
Server depreciation	875,000$(31.25$)	Constant
Server opex (Service/repairs)	43,750$ (1.56$)	Constant
PUE overhead	55,467$ (1.98$)	Constant
Utility monthly energy cost	252,179$ (9.01$)	Constant
Utility monthly power cost	117,600$ (4.20$)	Decreasing
Total	2,152,961$(76.89$)	Decreasing

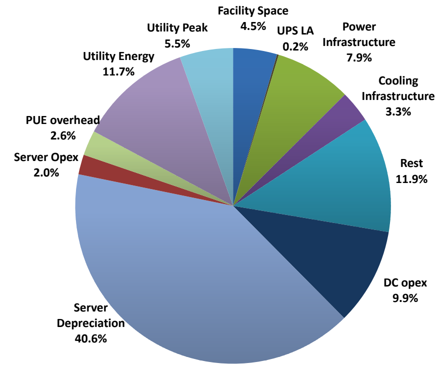


Figure 2: Total Cost of Ownership (TCO) [3]. TCO/server decreases as we increase servers under same power budget

posal over-provisions batteries and increases the cost of the distributed UPS. In return, we amortize the cost of several large components over a larger set of servers. The full TCO model described here can be found in [22].

4 The benefits of power oversubscription

In this section we evaluate the benefit of power oversubscription ignoring the cost of the power management solution that prevents power budget violations. This extra cost can be expressed in performance loss and hence income reduction for example when DVFS or consolidation is applied. Or additional cost for hardware modifications that ensure zero performance degradation, such as battery enabled power capping.

The three pie charts in figure 3 show the benefits of oversubscribing the supporting equipment at different levels of the power hierarchy. Power oversubscription at higher levels incurs extra costs. Oversubscription at the rack level, assuming there is sufficient rack space, comes effectively for free. At the PDU level we account for the extra rack cost as well as the additional facility space that the extra racks occupy. Finally at the cluster level on top of the rack and facility cost we need to account for the additional PDU cost to accommodate the newly added racks. As a result, the benefits in terms of TCO per server decrease as we oversubscribe higher in the power hierarchy. Note here that oversubscribing at lower levels is more beneficial, however at higher levels of the power hierarchy averaging effects of server power result in larger margins for power reduction and oversubscription, at no performance cost.

At this point we should stress that these benefits will vary significantly according to the underlying assumptions for power oversubscription as well as the component costs. To highlight this effect we perform a sensitivity analysis on the cost and the peak power of the server in figure 4. Power oversubscription becomes more effective with high power, low cost servers and less effective with low power, high cost servers. With smaller peak power servers we can pack more servers under the same supporting equipment and amortize capital expenses better. This means that the supporting equipment constitutes a smaller portion of the

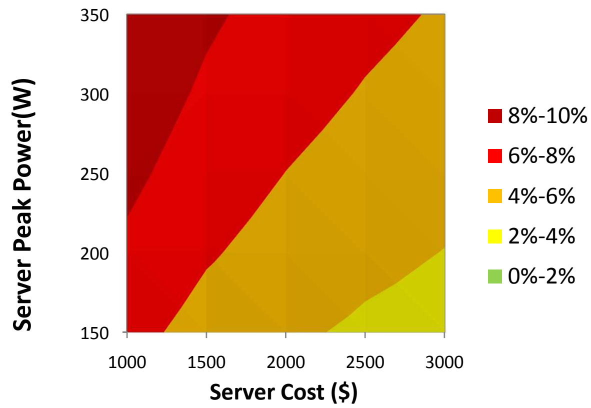


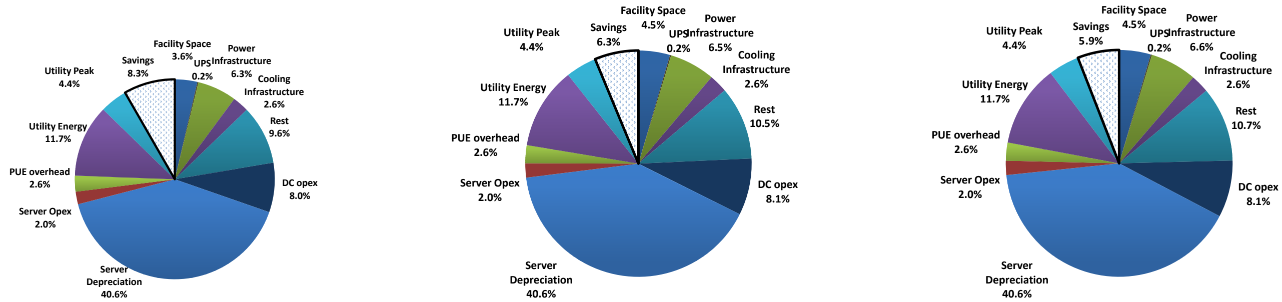
Figure 4: TCO per server reduction of placing 24% more servers under the same power infrastructure as the cost of the server and its peak power changes.

TCO per server to begin with and therefore power oversubscription is less helpful. Similarly, higher server cost for server with same peak power, translates to higher cost for the same number of server. This is equivalent to smaller portion of the TCO part for the supporting equipment and less effective power oversubscription. The best scenario in terms of power oversubscription savings is low cost, high power servers. For this scenario, we reduce the total cost of ownership per server by almost 10%. This reduction is achieved with 24% more servers under the same infrastructure. Such degree of oversubscription is possible with the UPS batteries that we describe in the next section.

5 Characterizing distributed UPS batteries

Current UPS designs rely on lead-acid batteries because of their ability to provide large currents for high power applications at low cost. In this section, we discuss alternative battery technologies for distributed UPSs, model battery behavior when employed for peak power capping, and elaborate on the selection of parameters (capacity, cost, depth of discharge) to minimize TCO/server.

The spider graph in Figure 5 compares the major com-



(a) Rack level (b) PDU level (c) Cluster level

Figure 3: The power oversubscription benefits at different levels of the power hierarchy. Oversubscribing at lower levels results in more savings.

peting battery technologies for high power applications, typical for servers, at the range of 12V and 15-30A: lead-acid (LA), Lithium Cobalt Oxide (LCO), and Lithium Iron Phosphate (LFP). Other technologies like NiCd, NiMH, or other lithium derivatives are excluded because they are dominated by one of the discussed technologies across all metrics. LA never performs best along any dimension except at low temperatures. While LA is cheapest per Wh, LFP offers an order of magnitude more recharge cycles, hence provides better \$/Wh/cycle than LA. LCO is the most expensive technology and provides comparable recharge cycles to LA. The advantage of LCO technology is its high volumetric density (Wh/l) and gravimetric density (Wh/Kg). Lithium batteries have longer service life than LA and also recharge faster. LFP has higher margins for over-charging and is safer than LA (may release toxic gases when over-charged) and LCO (may catch fire).

Properly selecting the technology and battery size depends on its use. UPS batteries in modern data centers are discharged only during a power outage. According to [29], the number of utility outages that affect data centers ranges from 1.5 to 4.4 per year. Therefore, cost, service life, and size are the most important parameters. The selection criteria become quite different when we re-purpose the batteries to be aggressively charged and discharged. Recharging cycles become crucial because continuous battery use may drastically shorten battery lifetime, resulting in frequent replacement costs that negatively affect TCO/server. Hence \$/Wh/cycle is a better metric than \$/Wh alone. Since LCO does poorly on both cost and cycles, it is not considered further.

We now focus on the per server distributed UPS design and explore the degree of overprovisioning that is most financially beneficial. Battery cost is estimated based on its 20h-rated capacity in Amp-hours (Ah) and the cost per Ah. We derive the required battery capacity based on the amount of power we want to shave and the corresponding energy stored in a battery for a given daily power profile. We derive the cost per Ah from [1, 8]. Tables 2 and 3 show all the inputs for the battery sizing estimation.

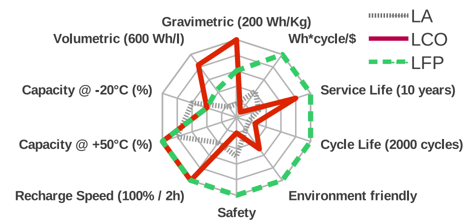


Figure 5: Comparison of battery technologies [1, 40]

Input	Value LA	Value LFP	Reference
Service time	4yrs	10yrs	[41, 8]
Battery Cost per Ah	2\$/Ah	5\$/Ah	[8, 1]
Depth of Discharge	40%	60%	Estimated (see figure 8)
Peukert's exponent	1.15	1.05	[18]
Existing Server Bat. Capacity	3.2Ah		[13]
Recharge Cycles	f(DoD) – Table 3		[38, 41]
Battery Voltage	12V		[13]
Max Bat. Discharge Current	23A		Estimated (ServerPeak * PSUeff / Voltage)
PSUeff	0.8		[4]
Discharges per day	1		Based on data from [12]
Battery losses	5%		[35, 42]

Table 2: Input values for battery cost estimation.

To derive the required battery capacity, we first set a peak power reduction goal and estimate the total energy that needs to be shaved at the data center level over the period of a day. We assume all batteries get charged and discharged once per day because, according to [12], all the traffic profiles of large distributed applications demonstrate a single peak. The daily shaved energy is equivalent to the integral between the power curve and the flat power line we set as the peak goal. For simplicity we model a power peak as a diurnal square pulse with a specified height and duration. For that workload, the required data center discharge energy is given by equation 2.

$$E_{DataCenter} = DataCenterPeakPower \times PowerReduction \times PeakTimePerDay \quad (2)$$

For our analysis, we assume that all servers discharge their batteries at the same rate and there is no day-to-day variation in energy demand. Equation 3 estimates the

DoD (%)	10	20	30	40	50	60	70	80	90	100
Rcycles LA	5000	2800	1860	1300	1000	830	650	500	410	330
Rcycles LFP	100000	40000	10000	6000	4000	3000	2000	1700	1200	1000

Table 3: Recharge cycles as a function of depth of discharge (DoD). Deep battery discharge results in a fewer recharge cycles[38, 41].

energy that each battery should provide to the associated server. Since the distributed battery is attached after the power supply the power drawn from the battery goes directly to the motherboard and is not wasted on losses of the Power Supply Unit (PSUefficiency).

$$E_{server} = \frac{E_{DataCenter} \times PSUefficiency}{N_{servers}} \quad (3)$$

Given the energy each battery must provide, we estimate the energy stored per battery and the corresponding battery capacity using Peukert's law. This relation is given by equation 4, where C_{1h} is the battery capacity in Ah (1h means that the battery capacity, equivalent to charge, is measured drawing constant current for 1h), I is the discharge current, PE is Peukert's exponent, and T is the battery discharge time [37, 35]. Lead-acid batteries typically have a Peukert's exponent in the range of 1.05-1.25 while Lithium Iron Phosphate batteries are in the range of 1.03-1.07 [18].

$$T = \frac{C_{1h}}{I^{PE}} \Rightarrow C_{1h} = T \times I^{PE} = \frac{E_{server}}{V \times I} \times I^{PE} = \frac{E_{server}}{V} \times I^{PE-1} \quad (4)$$

We also account for battery depth of discharge (DoD), the degree to which we allow the battery to be drained. Fully discharging the battery (100% DoD) to extract the required amount of energy would seriously degrade the lifetime of the battery and translate to higher battery replacement costs (see table 3). Limiting DoD also allows us to use excess capacity for power capping without increasing exposure to power failures. Consequently, we only want to discharge the battery partially. However, the less we discharge a battery, the larger battery capacity we need in order to discharge the same amount of energy. For discharge current, we conservatively assume the max value of the server current ($I_{MAX} = 23A$). Additionally, batteries lose a portion of their capacity as they age. Once they reach 80% of their original capacity, battery manufacturers consider them dead. We pessimistically take this effect into account by scaling the capacity by a factor of 1/0.8. Using equation 4, we get the provisioned 1h-rated battery capacity for each server battery (equation 5).

$$C_{1h}\ prov. = C_{1h} \times \frac{1}{DoD} \times \frac{1}{0.8} = \frac{E_{server}}{V} \times I_{discharge}^{PE-1} \times \frac{1}{DoD} \times \frac{1}{0.8} \quad (5)$$

Finally, we convert the 1h-rated capacity into 20h-rated capacity [35, 37], the value reported by battery manufacturers.

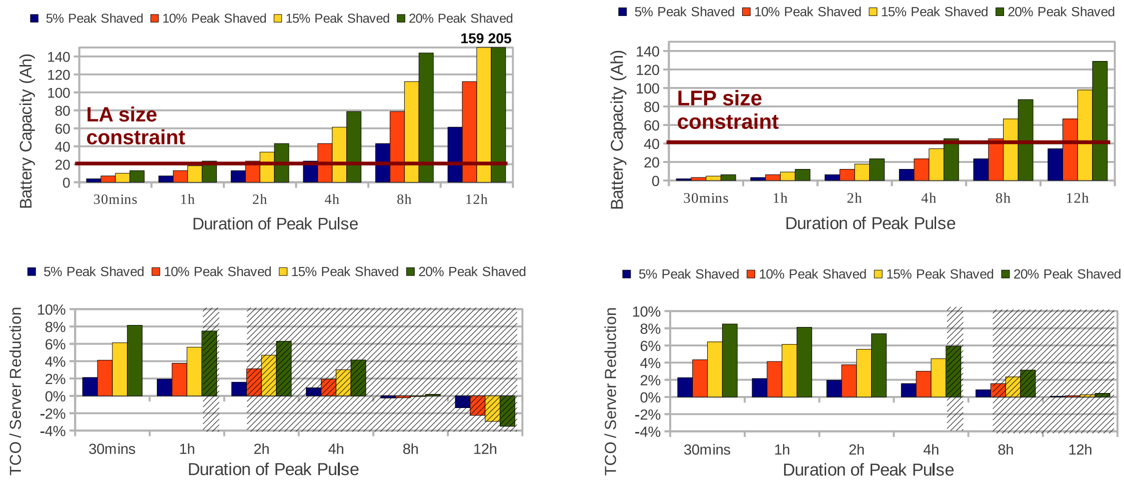
The previous capacity estimation methodology allows us to translate a peak power reduction goal to per-server provisioned battery capacity and the associated cost. To compute the monthly UPS depreciation, we also need to know the average battery lifetime. The battery lifetime is equal to the min of the battery service time in months and the number of recharge cycles as a function of depth of discharge, divided by 30 (one recharge cycle per day):

$$UPSDepr = \frac{C_{20h}\ prov. \times BatteryCostPerAh \times N_{servers}}{MIN\left(serviceLife, cycles(DoD)/30\right)} \quad (6)$$

We use the described equations to contrast LA with LFP technologies as we vary the peak time in the power profile, study the effect of decreasing battery cost per Ah, and identify the depth of discharge that minimizes TCO/server. Figure 6 shows the provisioned battery capacity for a given peak power time and a targeted reduction in peak power as well as the respective TCO/server reduction with power oversubscription at the rack level. More energy needs to be stored in the battery to achieve the same reduction in peak power as the duration of peak power demand increases. Hence, the cost of the distributed UPS increases. In the LA case, over-provisioning is no longer helpful when the peak power lasts for 12 hours. This means that the additional distributed UPS cost is greater than the reduction of TCO/server due to amortization of the infrastructure costs on more servers. LFP batteries remain marginally beneficial at 12 hours of peak demand. Size constraints only allow shaving 5% of the 2-hour peak demand, in the LA case, while we can shave 5% of an 8-hour pulse with LFP. In the TCO/server diagrams in figure 6, we denote the battery capacities that do not fit in a 2U server by hatch shading the respective columns. For the same spike duration, it always makes sense to shave more peak with a bigger battery, within size limitations. To further quantify these profits we find, using the analysis of section 3, that 6.8% monthly TCO/server reduction translates to \$6.4 per month per server, or more than \$21M over the 10-year lifetime of a data center with 28,000 servers.

Figure 7 presents the monthly TCO/server savings as the battery costs change. The projection for LA batteries is that costs will not change, while LFP prices are expected to decline due to the push for cheaper hybrid and electric cars [7]. For these graphs we assume that LFP cost reduces yearly at 8% [2]. At 4h peak per day, we achieve 7% TCO/server reduction for lead-acid, ignoring space considerations, while this value drops to 1.35% for a battery that fits within a 2U server design. Using LFP batteries today we can achieve 8.5% TCO/server reduction and these savings will increase to 9.6% in the next 6 years.

Figure 8 shows the relation between depth of discharge and the TCO/server gains for both LA and LFP technology. There is a clear peak for the values 40% and 60% DoD, respectively. For low DoD values, the battery costs dominate the savings, because we need larger batteries to provide the same capping. For large DoD values, the lifetime of the battery decreases and more frequent replacements increase the UPS cost. The peak reduction of TCO/server occurs when the number of recharge cycles / 30 is equal to the battery service life. Note that due to the battery overprovisioning, less



(a) LA (b) LFP

Figure 6: Battery capacities for different pulse widths and portion of peak power shaved. We also show the monthly TCO per server savings, assuming current battery costs, for the specified capacities of Lead-acid (LA) and Lithium Iron Phosphate (LFP) batteries. When the battery cannot fit within a 2U server, the associated savings are hatch shaded.

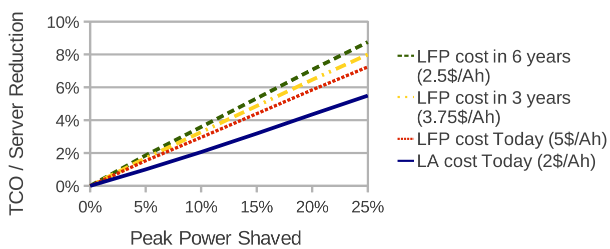
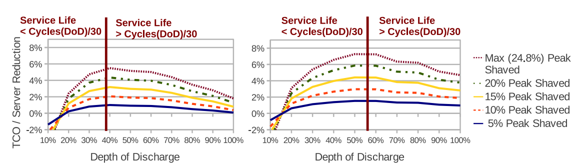


Figure 7: For the 2h pulse we show the projection of savings (ignoring space constraints) as the battery cost changes in the future [2].

than 5% charge can sustain the server for 1 min and ensure data center continuity. Therefore, battery lifetime considerations affect TCO/server well before data center continuity becomes a concern.

To summarize our discussion on battery technologies and battery properties, we conclude: (1) Battery-based peak power shaving using existing batteries is only effective for brief spikes. To tolerate long spikes, larger batteries are necessary. However, the benefits from increased peak power shaving outweigh the extra battery costs even when high demand lasts 12 hours. (2) LFP is a better, more profitable choice than LA for frequent discharge/recharge cycles on distributed UPS designs. This is due to the increased number of cycles and longer service lifetime, better discharge characteristics, higher energy density, and the reduction in battery costs expected in the near future. (3) It makes sense to increase the capacity of the battery to the extent that it fits under the space constraints. This translates to increased power reduction and more savings. (4) For each battery technology, there is a depth of discharge value that maximizes savings (40% for LA and 60% for LFP). This is the point where battery lifetime is no longer limited by the battery service time and needs to be replaced earlier due to frequent charging and discharging.



(a) LA (b) LFP

Figure 8: The relation between targeted depth of discharge and the reduction in TCO.

6 Related Work

Peak Power Provisioning and Capping: Reducing power consumption in server clusters is a well-studied problem in the literature [34, 26, 24, 14]. The overall idea is to combine CPU throttling, dynamic voltage/frequency scaling (DVFS), and switching entire servers on/off depending on the workload. Raghavendra, et al. [34] note that more efficient power management solutions are possible by managing power at the rack level than at individual blades. They devise proactive and reactive policies based on DVFS to cap power budgets at the rack level. Nathuji and Schwan [26] introduce the notion of power tokens to deal with heterogeneity across hardware platforms. Govindan, et al. [14] combine applications with heterogeneous characteristics in terms of burstiness. As a result, the power budget is exceeded statistically infrequently. DVFS is used as a failsafe mechanism to prevent against lasting peak power violations.

Femal et al. [11] were among the first to use formal control theory to maximize throughput while capping power. Raghavendra, et al. [31] extend the control theory idea to present a hierarchy of coordinated controllers that cap power across different levels of the power hierarchy and minimize performance impact. They argue for nesting controllers that operate at different time granularities to ensure stability and emphasize the information flow between the controllers.

Using batteries in data centers: Battery power management has been studied in the embedded/mobile system domain with various works proposing techniques to adjust the drain rate of batteries in order to elongate the system operation time [27, 33, 32, 36]. Prior research has also investigated analytical models for battery capacity and voltage in portable devices [27, 36, 20]. Govindan, et al [15] introduce the idea of reducing data center peak power by leveraging the stored energy in a centralized UPS. During peak load, power from the UPS batteries augments the main grid, effectively hiding the peak from the utility service. During low load, the batteries recharge, consuming additional power.

In a follow-up work [17], they extend their prior work to also use distributed UPSs for peak power capping. That work focuses on power capping at the rack, using small lead-acid batteries to shave peak power. This approach allows them to prevent rare, brief power emergencies without performance degradation and relies on DVFS and load migration for peaks longer than several minutes. In our work, we examine solutions at multiple levels of the power hierarchy, show the financial advantages of more aggressive batteries with a more detailed model that incorporates battery lifetime considerations, and employ solutions that sacrifice no performance – the desired solution in a performance-sensitive data center under peak load.

In a separate work [16], the same authors also argue for a distributed UPS solution from a cost and reliability perspective. They find that a hybrid distributed UPS placement, at PDU and server level, yields the most promising topology. They do not consider battery energy for peak power capping in that work, but this finding provides additional motivation for our work on the use of distributed batteries for power capping.

7 Conclusions

State-of-the-art data centers such as Google's and Facebook's have adopted a distributed UPS topology in response to the high cost associated with a centralized UPS design. In this work we explore the potential of using battery-stored energy in a distributed UPS topology to shave peak power. We describe how to provision the capacity of the battery and elaborate on how recharge cycles, the depth of discharge, and the workload power profile affect the potential for peak power shaving. We find investing in bigger batteries worthwhile and show that aggressive UPS battery sizing creates higher power oversubscription margins, and further reduces TCO per server.

References

[1] AA Portable Power Corp. Portable Power Product design, assemble and quality control. `http://www.batteryspace.com/lifepo4cellspacks.aspx`.

[2] D. Anderson. An evaluation of current and future costs for lithium-ion batteries for use in electrified vehicle powertrains. Master's thesis, Duke University, 2009.

[3] APC. InfraStruxure Total Cost of Ownership. `http://www.apc.com/tools/isx/tco/`, 2008.

[4] Climate Savers Computing. Power supply efficiency specifications. `http://www.climatesaverscomputing.org/resources/certification`, 2011.

[5] Duke Energy. Utility bill. `http://www.duke-energy.com/pdfs/scscheduleopt.pdf`, 2009.

[6] Dupont Fabros Technology, Inc. Sec filing (s-11) 333-145294, August 9, 2007.

[7] Economist. In search of the perfect battery. `http://www.mitenergyclub.org/assets/2009/9/25/Economist_Batteries2008.pdf`, Mar. 2008.

[8] Electic motor sport. EV construction, thundersky batteries. `http://www.electricmotorsport.com/store/ems_ev_parts_batteries.php`, 2001.

[9] Facebook. Hacking conventional computing infrastructure. `http://opencompute.org/`, 2011.

[10] X. Fan, W.-D. Weber, and L. A. Barroso. Power provisioning for a warehouse-sized computer. In *ISCA*, 2007.

[11] M. E. Femal and V. W. Freeh. Boosting data center performance through non-uniform power allocation. In *ICAC*, 2005.

[12] Google. *http://www.google.com/transparencyreport/traffic/*.

[13] Google Summit. `http://www.google.com/corporate/datacenter/events/dc-summit-2009.html`, 2009.

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

[15] S. Govindan, A. Sivasubramaniam, and B. Urgaonkar. Benefits and limitations of tapping into stored energy for datacenters. In *ISCA*, 2011.

[16] S. Govindan, D. Wang, L. Chen, A. Sivasubramaniam, and A. Sivasubramaniam. Towards realizing a low cost and highly available datacenter power infrastructure. In *HotPower*, 2011.

[17] S. Govindan, D. Wang, A. Sivasubramaniam, and B. Urgaonkar. Leveraging stored energy for handling power emergencies in aggressively provisioned datacenters. In *ASPLOS*, 2012.

[18] F. Harvey. Listing of ev batteries. `http://www.39pw.us/car/batteryTable.html`, 2009.

[19] U. Hoelzle and L. A. Barroso. *The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines*. Morgan and Claypool Publishers, 2009.

[20] M. Jongerden and B. Haverkort. Battery modeling. Technical report, TR-CTIT-08-01, CTIT, 2008.

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

[22] V. Kontorinis. Battery-aware Data Center TCO models. `http://cseweb.ucsd.edu/~tullsen/DCmodeling.html`, 2012.

[23] H. Lim, A. Kansal, and J. Liu. Power budgeting for virtualized data centers. In *USENIX*, 2011.

[24] D. Meisner, B. Gold, and W. Thomas. Powernap: Eliminating server idle power. In *ASPLOS*, 2009.

[25] R. Nathuji and K. Schwan. Virtualpower: coordinated power management in virtualized enterprise systems. In *SOSP*, 2007.

[26] R. Nathuji and K. Schwan. Vpm tokens: virtual machine-aware power budgeting in datacenters. In *HPDC*, 2008.

[27] D. Panigrahi, S. Dey, R. R. Rao, K. Lahiri, C.-F. Chiasserini, and A. Raghunathan. Battery life estimation of mobile embedded systems. In *VLSI Design*, 2001.

[28] S. Pelley, D. Meisner, P. Zandevakili, T. F. Wenisch, and J. Underwood. Power routing: dynamic power provisioning in the data center. In *ASPLOS*, 2010.

[29] Ponemon Inst. *National Survey on Data Center Outages*, 2010.

[30] S. press release. Savvis sells asserts related to two datacenters for $200 million. `http://www.savvis.com/en-US/Company/News/Press/Pages/SAVVIS+Sells+Assets+Related+to+Two+Data+Centers+for+200+Million.aspx`, June 29 2007.

[31] R. Raghavendra, P. Ranganathan, V. Talwar, Z. Wang, and X. Zhu. No "power" struggles: coordinated multi-level power management for the data center. In *ASPLOS*, 2008.

[32] D. N. Rakhmatov. Battery voltage prediction for portable systems. In *ISCAS*, 2005.

[33] D. N. Rakhmatov and S. B. K. Vrudhula. Energy management for battery-powered embedded systems. *TECS*, 2003.

[34] P. Ranganathan, P. Leech, D. Irwin, and J. Chase. Ensemble-level power management for dense blade servers. In *International Symposium on Computer Architecture*, June 2006.

[35] T. B. Reddy and D. Linden. *Linden's Handbook of Batteries (4th edition)*. McGraw-Hill, 2011.

[36] P. Rong and M. Pedram. An analytical model for predicting the remaining battery capacity of lithium-ion batteries. In *DATE*, 2003.

[37] SmartGauge. Peukert's law equation. `http://www.smartgauge.co.uk/peukert.html`, 2011.

[38] M. Swierczynski, R. Teodorescu, and P. Rodriguez. Lifetime investigations of a lithium iron phosphate (LFP) battery system connected to a wind turbine for forecast improvement and output power gradient reduction. In *BatCon*, 2008.

[39] W. P. Turner and K. G. Brill. Cost Model: Dollars per kW plus Dollars per Square Floor of Computer Floor, Uptime Inst. White paper, 2009.

[40] B. University. Online university education about batteries. `http://batteryuniversity.com/`, 2003.

[41] Windsun. Lead-acid batteries: Lifetime vs Depth of discharge. `http://www.windsun.com/Batteries/Battery_FAQ.htm`, 2009.

[42] L.-L. Zhang, G. Liang, A. Ignatov, M. C. Croft, X.-Q. Xiong, I.-M. Hung, Y.-H. Huang, X.-L. Hu, W.-X. Zhang, and Y.-L. Peng. Effect of Vanadium Incorporation on Electrochemical Performance of LiFePO4 for Lithium-Ion Batteries. In *Journal of Physical Chemistry*, June 2011.