Enabling Distributed Generation Powered Sustainable High-Performance Data Center

Chao Li, Ruijin Zhou and Tao Li
Intelligent Design of Efficient Architectures Laboratory (IDEAL)
Department of Electrical and Computer Engineering, University of Florida
{chaol, zhourj}@ufl.edu, taoli@ece.ufl.edu

Abstract — *The necessity for capping carbon emission has significantly restricted the potential of modern data centers. For this matter, both industry and academia are proactively seeking opportunities on cross-layer power management schemes that could open a door for sustainable high-performance computing platform. In this paper we investigate an emerging trend in the IT industry: using promising onsite distributed generation (DG) techniques to provide premium clean energy to the computing load.*

We develop data center power demand shaping (PDS), a novel technique that allows data centers to utilize onsite green energy efficiently. In contrast to prior design, PDS takes advantage of a so-far unexplored power supply feature, i.e., the load following capabilities of DG systems to avoid the high performance penalty issue incurred during supply tracking. In addition, PDS features two adaptive power management schemes: DGR Boost and UPS Boost. These two workload-aware optimization methods leverage mature computer tuning knobs to achieve attractive data center performance improvement. Using real-world data center traces and industry data of distributed generation systems, we show that our technique can come within 1.2% performance of an ideal oracle, which is roughly a 37% improvement over existing supply tracking based design. Our design could save over 100 metric tons of carbon emissions annually for a 10MW data center.

1. Introduction

Computer system design inevitably enters the landscape of design for sustainability as the data center power footprint has become a global concern. In the past ten years, Google server's electricity demand has increased almost 20-fold [1]. The huge IT energy consumption not only increases the total cost of ownership (TCO) but also leaves profound impact on the environment. According to a McKensey Quarterly report, the annual CO_2 emissions of computing systems will research 1.54 metric gigatons within eight years, which could make IT company among the biggest greenhouse gas emitters by 2020 [2]. Consequently, renewable energy powered data centers are gaining growing popularity in both IT industry and academia as a way to tackle the dual challenges of energy shortage and environmental issues [3-9].

Existing proposals on renewable energy-aware power management schemes largely emphasize adapting the computer load to the time-varying power budget [3-9]. We broadly categorize these techniques into two types: 1) load tuning based design; and 2) job scheduling based design. While the former approach leverages performance scaling techniques (e.g., DVFS and server power state tuning) to track the time-varying renewable power budget [3-5], the later approach schedules job requests based on the renewable energy availability [6-9]. Since these techniques are driven by variable and intermittent power supply, they typically suffer extended job turnaround time. They can hardly maintain desired instantaneous throughput and service availability without substantial utility grid support.

In this paper we present a fundamentally different design, which allows near-oracle performance of computing systems on a variety of green energy resources. The key idea of our approach is that we put emphasis on enabling renewable energy supply to follow data center power demand, rather than forcing the IT load to track the variable power budget. To achieve this goal, we leverage distributed generation initiated by the smart grid technology [10].

Distributed generation (DG) refers to a variety of small, modular electric generators near the point of use. In recent years, DG has gained tremendous interest as an alternative source of power for IT industry. According to the U.S. Environmental Protection Agency (EPA), using DG in data center design could achieve great energy savings, significant environmental benefits, and high power reliability [11].

As shown in Figure 1, DG system encompasses a wide range of green energy technologies, such as photovoltaic module (PV), wind power, fuel cell, and bio-fuel based gas turbine. While PV/wind power depends on environmental condition, the outputs of fuel cells and gas turbines are tunable. They can provide a key supporting service called load following [12], which refers to the use of online generation equipment to track the changes in customer loads. Therefore, one can take advantage of the load following capabilities of these tunable DG systems to meet the time-varying IT power demand. Such design is non-trivial because it enables data center to run on renewable energy sources without compromising workload performance.

When employing distributed generation to build a better data center power provisioning architecture, challenges arise due to the unpredictable and fluctuating data center load. Figure 2 shows typical load following scenario that tracks customer load every 30 minutes. As can be seen, DG systems cannot provide fine-grained load demand following due to their limited response speed.

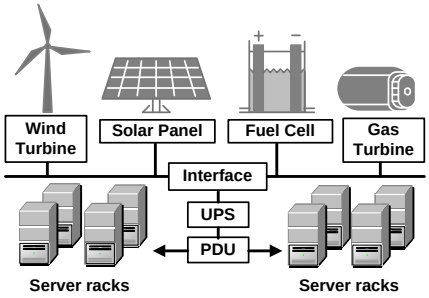


Figure 1: Distributed generation powered data center. The system can follow customer's load power demand

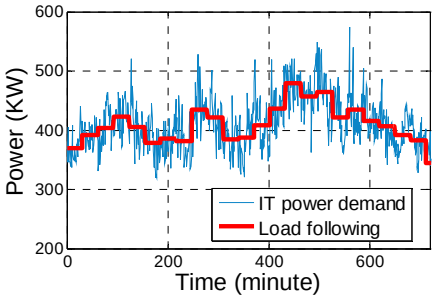


Figure 2: Load following scenario. Frequent IT load fluctuation hinders efficient load following

To handle the moment-to-moment load power demand, DG systems typically rely on large energy storage elements [12]. Such design not only increases the TCO (due to storage cost), but also incurs up to 25% roundtrip energy loss. More importantly, without careful power management, the disturbing load can cause frequent and excessive battery discharging activities, which may degrade the lifetime of these expensive electrical elements and quickly deplete the stored energy that is crucial for handling emergencies.

We propose data center power demand shaping (PDS), a novel power management approach that enables high-performance low-overhead data center operation on pure renewable energy sources. The novelty of PDS is two-fold. First, PDS intelligently trims data center load power and enables DG systems to follow the power demand efficiently. Second, PDS features two adaptive load tuning schemes that could boost data center performance and enable near-oracle operation during power demand trimming process. As a cross-layer power optimization scheme, our power management module resides between front-end distributed generation and back-end computing facilities to provide a coordinated tuning between the supply and load.

This paper makes the following contributions:

- We propose sustainable data center design that leverages onsite distributed generators. We characterize *load following*, a so-far unexplored power provisioning scheme in data centers. We show that a well designed load power management scheme could help data center operators to achieve the optimal benefits of load following. This means 8% additional energy utilization improvement, 1.3X battery lifetime increase, and significant savings in energy storage cost.
- We propose power demand shaping (PDS) mechanism for optimizing load following efficiency and workload performance in DG-powered data centers. Overall, PDS achieves 98.8% performance of an ideal oracle design. Compared to recent supply tracking based approach, PDS could improve the job turnaround time by 37% on average. To achieve the same performance as PDS, the state-of-the-art green data centers have to rely on utility power for about 50% of its runtime. Furthermore, PDS minimizes the overhead on energy storage devices, and improves the battery lifespan by up to 26%.

The rest of this paper is organized as follows. Section 2 introduces DG and microgrid. Section 3 characterizes load following. Section 4 proposes power demand shaping mechanism and our adaptive load tuning schemes. Section 5 describes experimental methodologies. Section 6 presents our evaluation results. Section 7 discusses related work and Section 8 concludes this paper.

2. Background

The IT industry is actively looking for opportunities in non-conventional power provisioning solutions such as distributed generation. For example, Apple [13] and Microsoft [14] have built their data centers that incorporate fuel cell technology; eBay [15] also plans to install 30 large-scale fuel cells which will provide 6MW power to its data center. HP [3] recently considers using bio-fuel based gas turbine in its Net-Zero data center. In this section, we introduce distributed generation technologies and discuss their impacts on data center design and operation.

2.1 Distributed Generation and Micro-grid

Distributed generation (DG) [10, 11] is an emerging trend of generating power locally to provide reliable, secure, and sustainable electrical energy to its consumers. Distributed generation encompasses several promising clean energy technologies such as gas turbine, biomass power, and fuel cell. These non-conventional power generators are known as *microsources* or *distributed energy resources* (DERs) which are modular units of small capacity (typically between several kilowatts to tens of megawatts) [16].

In order to harness clean energy from DERs, microgrid is proposed as a local electricity distribution network that focuses on flexible and intelligent management of DG systems. The responsibility of microgrid is to dynamically control the power flow in response to any disturbance and load changes. Although microgrid can import/export power from/to the utility power line, it is usually the last resort due to the low transmission efficiency, high peak power cost, and sustainability consideration [12].

In this paper we focus our attention on minimizing data center's reliance on conventional utility grid for reducing carbon footprint. Such design is also often preferable when the utility power is less reliable in some developing

countries. Further, for data centers that are built in remote areas, a grid-dependent operation can help to eliminate the expensive utility power line extensions (around $10k per km) [16].

2.2 Load Matching Challenge

Different from conventional bulk grid which has large amount of capacity inertia, distributed generation does not have reserved capacity [11]. The supply and storage of energy must be planned carefully in microgrid to ensure instantaneous demand and long-term energy balance [10].

In Table 1 we show the response speed of typical DG systems. Most energy storage devices have very fast response speed that could release power almost immediately (in ms level). As a result, they are widely used to handle moment-to-moment load oscillation and disturbances, which is referred to as *regulation* [12]. In contrast, gas turbines and fuel cells are typically too slow to meet the load power variation since the change of the engine speed or the chemical reaction in the fuel requires time. Therefore, they are used to track the intra- and inter-hour changes in customer loads, which is referred to as *load following* [12].

Although there is no strict rule to define the temporal boundary between regulation and load following, typically load following occurs every 10~15 minutes or more. It is not economically feasible to frequently adjust the output of distributed generators due to the increased performance cost and decreased fuel utilization efficiency.

The energy balance issue arises due to the fluctuating load in data centers and other computerized environments. Dynamic power tuning techniques (e.g., DVFS), frequent on/off power cycles, stochastic user requests, and data migration activities can cause varying degree of load power variation. Since distributed generators are generally placed near or at the point of energy consumption, they are often exposed to the full fluctuation of local IT loads rather than experiencing the averaging effect seen by larger, centralized power system. As a result, energy balancing becomes rather difficult in distributed generation powered data centers.

DG Systems	Response Speed	Startup Time
Lead-acid battery	Immediate	N/A
Flywheel	Immediate	N/A
Fuel cell	30 sec ~ 5 min	20~50 min
Gas turbine	10s of seconds	2~10 min

Table 1: Response speed of DG systems [17,18]

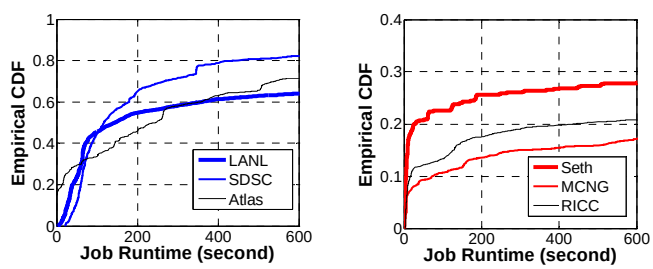
3. Analyzing Load Following in Data Centers

In this section we characterize load following in data centers using real-world HPC workload traces and industry data of distributed generation systems (detailed in Section 5). We show that enabling load following in data centers is a challenging task. Conventional load following scheme results in sub-optimal energy utilization, low energy storage lifetime, and poor cost-effectiveness.

3.1 Energy Utilization

The energy utilization problem arises as typical microsources are not able to respond rapidly to the frequent IT load variation. In Figure 3 we show the cumulative distribution functions (CDFs) of job runtime in HPC data centers. We consider both short-running workload (average job runtime < 1h) and long-running workload (average job runtime > 1h) traces. For short-running workload traces, we observe that more than 40% submitted jobs are finished with 100 seconds. For long-running workload traces, nearly 30% of the submitted jobs show short runtime that does not exceed 10 minutes. Most of these short jobs will be granted with individual computing nodes, resulting in transient load power demand fluctuation.

Workload fluctuation can be the major obstacle of achieving high renewable energy utilization in a distributed generation powered data center. Figure 4 shows the energy utilization of data center with varying load following intervals. We dynamically adjust the DG power output based on the peak power demand of the last load following period. For a 15-min load following, the energy utilization is 97% on average. For a 60-min load following, the energy utilization drops to 90% on average. Apparently, the energy efficiency decreases as load following becomes coarse-grained. However, even if we perform load following on an hourly basis, the energy utilization still outperforms over-provisioning (Over-P) based design which has an average energy utilization of 66%.



(a) Short-running traces **(b) Long-running traces**

Figure 3: CDF of workload runtime in a zoomed view. Short jobs may result in fine-grained load fluctuations

3.2 Energy Storage Lifespan

Batteries are critical components in both distributed generation systems and uninterruptable power supply (UPS) systems. These energy storage devices are prone to electrical wear out under irregular charging and discharging regime [19]. Throughout the history of battery discharge event, two factors affect the battery failure rate significantly. The first is the depth of discharge (DOD), which indicates how much stored energy has been used (0% DOD = full capacity, 100% DOD = empty). The second factor is the discharge current which represents the rate of discharging. The battery lifetime will be decreased whenever the battery cell is discharged at a faster rate than the rated rate [20].

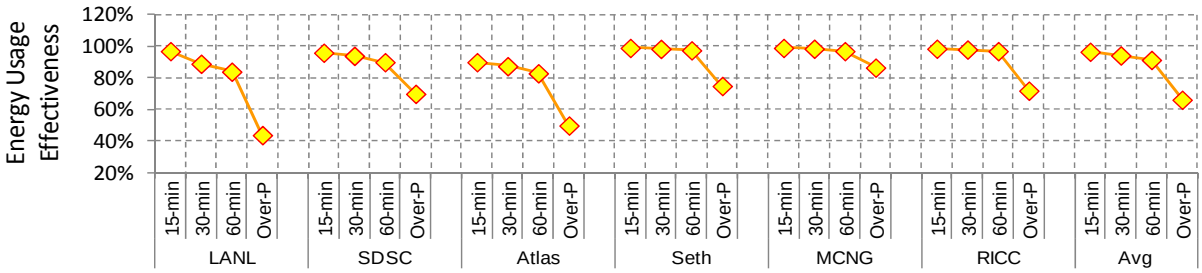


Figure 4: The energy utilization decreases when load following becomes coarse-grained. Over-provisioning based design (Over-P) is the worst-case which always generates the peak power (no load following capability)

In this study we use a battery lifetime evaluation method that captures the aforementioned two primary determinants of battery failure [20]. This technique predicts the battery lifespan based on manufacture's battery performance data and the measured battery usage record of duration T. Assume the rated charge life (in Ah) is L_R. The estimated lifetime is given by:

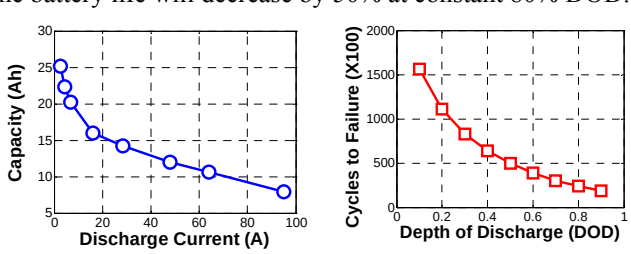
$$L = \frac{TL_R}{\sum_i d_i^{eff}} = \frac{TL_R}{\sum A_i B_i d_i^{act}} \quad , \tag{3.1}$$

where d^{eff} is the effective discharge of a single discharge event, d^{act} is the measured actual discharge, A_i is the scaling factor that represents the capacity degradation effect under high discharging rate, and B_i is the scaling factor that represents the lifetime degradation effect under high DOD. Both scaling factors are calculated from the best-fit functions of the manufacture's data sheets:

$$A_i = f_1(C_R, I_d) \tag{3.2}$$

$$B_i = f_2(D_R / D_A) \tag{3.3}$$

In equations 2 and 3, A_i is a function of the rated amp-hour capacity (C_R) and the actual discharging current (I_d); B_i is a function of the rated depth of discharge (D_R) and the measured actual depth of discharge (D_A). We model a 12V 6-cell valve-regulated lead-acid battery (VRLA), which is widely used in today's UPS systems. Figure 5 shows the battery behavior under varying discharging current and DOD. While the rated capacity is 24Ah at a 20-hour discharging rate, the capacity drops to only 12Ah at a 15-min discharging rate (15 min is the typical UPS ride-though duration). Compared to the 40% constant DOD operation, the battery life will decrease by 50% at constant 80% DOD.



(a) Capacity vs dischage rate **(b) Cycle life vs DOD**

Figure 5: Performance data of VRLA battery. High DOD and discharge rate will shorten the battery lifespan

In Figure 6, the load following policy is to adjust the distributed generation output based on the mean power demand of the last period. The minimal load following interval is 5-min and all the results at other interval values are normalized to it. Since the actual granularity of load following depends on several factors (e.g., the generator characteristic, control latency, and operation policy), the battery lifespan can vary significantly. If the best achievable (or economically feasible) load following interval is 30-min, the system will incur 57% battery lifetime degradation. This also implies that if the load power demand becomes less fluctuating, we can achieve 1.3X lifetime improvement.

3.3 Cost Analysis

Depending upon the load following schemes, the total cost of ownership (TCO) may vary. The main reason is that each load following scheme requires different amount of onsite battery capacity. Poor load following management and bursty load power demand increase the burden of regulation significantly, and therefore require high-capacity energy storage devices.

In Figure 7 we show the amortized capital expenditure (CapEx) of batteries under varying capacities and load following intervals. We assume a 10MW data center scale for both short-running and long-running workloads. For each load tuning interval value, we calculate the minimum required battery capacity (i.e., 1X) that can safely meet the load power regulation requirement. We then scale up the battery installation capacity (up to 2X) to perform a sensitivity analysis. We calculate the amortized cost over the battery's lifetime. As shown in Figure 7, small capacity does not necessarily mean cost saving since each battery cell in this case experiences more regulation events and fails quickly. Compared to battery capacity scaling, load following shows greater influence on the CapEx and results in up to 50% cost savings on batteries.

Summary: Due to data center load power fluctuation, peak power based load following incurs energy cost while mean power based load following incurs battery capacity cost. Batteries are crucial energy regulation components but incur up to 57% lifetime degradation under fluctuating IT load. To improve overall design efficiency, we must intelligently integrate load following into data center operation.

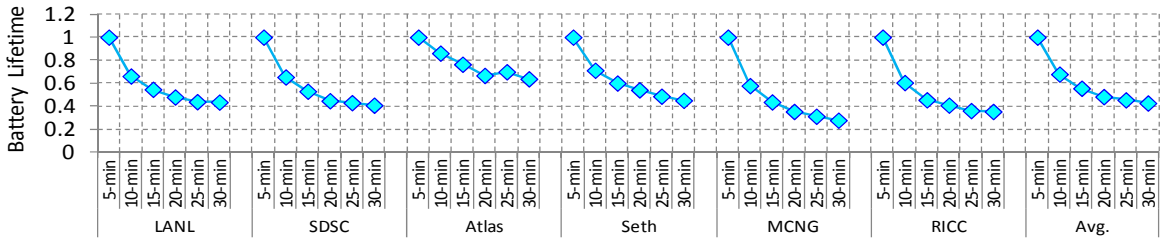
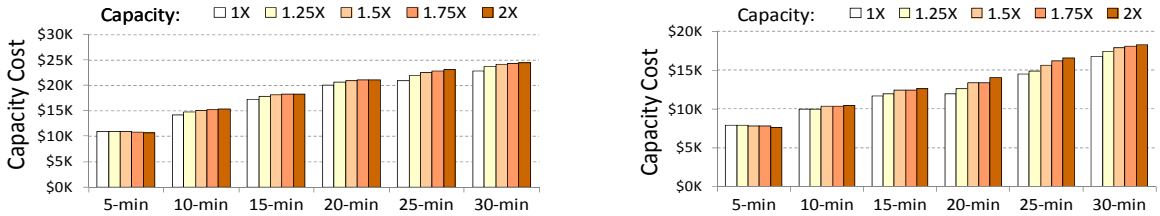


Figure 6: Normalized battery lifespan under different load following intervals. Conventional load following schemes cannot handle the bursty load and therefore incur varying degree of battery wear-out



(a) LANL (short-running workload) **(b) Seth (long-running workload)**

Figure 7: Battery cost under different load following intervals and capacity. Due to frequent power demand fluctuation, small battery capacity does not show the cost advantage it should have

4. Enabling Distributed Generation Powered Sustainable High-Performance Computing

In this section we propose power demand shaping mechanism to enable efficient load following in distributed generation systems. Our design overcomes many drawbacks of conventional load following such as low energy utilization, large battery requirement, and short battery lifespan. We first introduce the architecture of distributed generation powered data centers. Afterwards, we will demonstrate our adaptive load tuning scheme that intelligently coordinates DG system and the IT load to allow high-performance sustainable computing.

4.1 Architecture Overview

Figure 8 depicts our distributed generation powered data center. We adopt typical microgrid power provisioning hierarchy for managing various distributed energy resources. The design consists of a group of radical feeders to provide power to electrical loads. Microsources are connected to the feeder through circuit breakers and appropriate power electronic interfaces (PEI). All the circuit breakers and power electronic interfaces are coordinated by the microgrid energy management module (EMM), which provides real-time monitoring and autonomic control of the power generation [21]. Although microgrid can connect to the utility through a single point called point of common coupling (PCC), we focus on standalone microgrid due to sustainability and cost-effectiveness reasons. Such islanded mode is also a remarkable feature of microgrid to avoid power quality issue in the main grid [10, 21].

The connected electrical loads may be critical or non-critical. HPC servers are typically sensitive and mission-critical loads that require stringent power quality and sufficient power budget. We assign these loads with controllable and stable microsources such as gas turbines and fuel cells. On the other hand, non-critical loads (such as normal data processing machines) can be curtailed without affecting customer benefits significantly. In Figure 8, the non-critical loads are powered by intermittent renewable power supplies. These loads are flexible enough to be lowered or shed as per necessity when power generation is not sufficient. In this study we are primarily interested in leveraging controllable microsources to follow the demand of critical data center load. In Figure 8, each server cluster is connected to the power distribution unit (PDU) via redundant power delivery paths to ensure uninterrupted operation in the event of PDU failure.

We use a power demand controller (PDC) to manage the server clusters. It keeps monitoring the job running status and the overall cluster power demand. It also coordinates with the job scheduler for the purpose of better computing resource allocation. Meanwhile, the controller communicates with the microgrid EMM to obtain current distributed generation levels and dynamically inform the EMM for necessary generation policy adjustment. The microgrid typically uses frequency droop control to monitor and adjust the on-site power generation [22]. This scheme uses a small change in the power bus frequency to determine the amount of power that each generator should put into the power bus.

Microgrid typically includes energy storage devices to provide necessary startup time for generators' output to ramp up when system changes. In this study we leverage UPS system to provide necessary stored energy, as shown in Figure 8. Note that although our proposed technique also applies to centralized UPS, we choose distributed UPS system which shows higher efficiency and better power management capability [22].

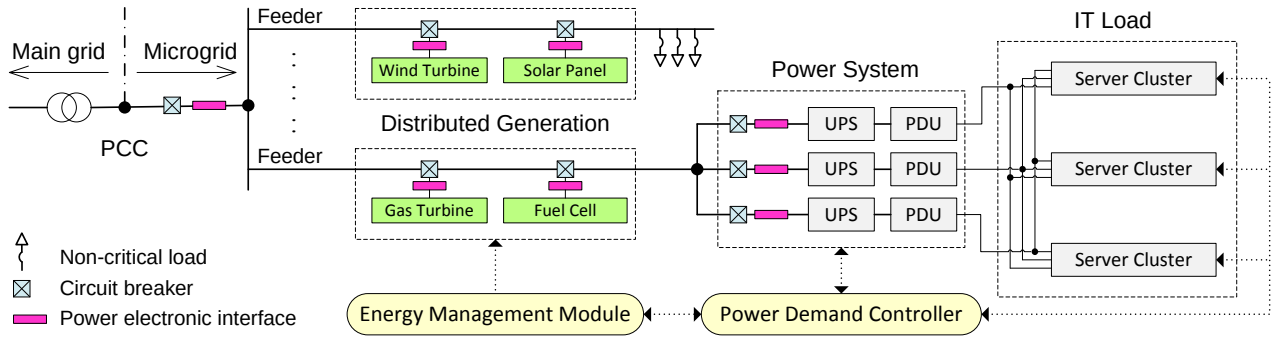


Figure 8: Architecture of distributed generation powered data center

4.2 Power Demand Shaping Mechanism

In contrast to prior art which primarily focus on harvesting the time-varying renewable energy, we observe that one can leverage the load following capability of distributed generation to jointly achieve high-performance and sustainability. In this paper, we propose power demand shaping (PDS) technique to better utilize onsite distributed generation in modern HPC data centers. The idea is to aptly influence IT power demand to match the planned power usage curve that yields high load following effectiveness.

While one can transform the IT power demand into many different shapes, we found that a square-wave-like demand curve is the most convenient, as shown in Figure 9. In this case, the distributed generation system only needs to increase/decrease its generation at each end of the load tuning cycle (e.g., every 15 minutes) to meet the new power budget goal. Since the load power demand becomes less bursty and fluctuating after tuning, there is no need to install bulky energy storage devices onsite. In this paper we mainly leverage the UPS to handle the ramping period and the temporary power discrepancy between supply and load. By eliminating the unnecessary energy storage capacity, our design also becomes more cost-effective and sustainable.

Figure 9 demonstrates our power demand shaping mechanism in terms of power supply behavior and load performance status. The proposed scheme uses a two-step control in which data center load is adaptively managed and the distributed energy is carefully scheduled. The two steps are further described below:

Step 1: Maintain constant power demand

During each control period, the PDS controller uses power capping to maintain a constant power demand. Assume the original load power demand (i.e., the expected power consumption without any load tuning) is L; we scale down the load performance whenever L exceeds the power demand goal D. If at certain point L drops below D, the data center will increase its load performance, and thereby revert to its original power demand shape.

To maintain a constant power demand, we need to dynamically tune the node performance level when data center load varies. There are various scheduling schemes. In this paper we adjust the node frequency in a round-robin manner for simplicity and fairness. In addition, one can use a priority-based scheme. It associates each application with a priority and makes a scheduling choice based on the priorities. For example, a group of nodes with high priority gets higher frequency.

Depending on the system's hardware characteristics and BIOS support, the actual performance scaling policy may vary. In this paper we take advantage of performance states (a.k.a. P-states) to dynamically adjust CPU frequency to match the required budget. This is a technique to manage power by running the processor at a less-than-maximum clock frequency. We do not use voltage scaling because it may not yield noticeable cubic power saving in the real measurement [24]. It has been shown that the power-to-frequency curve for both DFS and DVFS can be approximated as a linear function [24]. As a result, we use only frequency scaling in our design and optimization for simplicity and practicality.

Step 2: Re-schedule DG generation

In contrast to existing power capping scenario, PDS deals with time-varying power capping budget. At the end of each control period, the controller will inform the distributed generation EMM to adjust the generation level for the next load following period.

To calculate the amount of generation adjustment, PDS calculates the mean value of the original power demand in the past control period. The actual DG power adjustment is selected from the mean value and current demand measurement, depending on whichever is greater. After that, the distributed generation will gradually increase its output and we assume a 1-min power ramping duration for the generators. During this period, we leverage onsite UPS storage to provide necessary power support.

In addition to the scheduled generation adjustment, the controller also assigns a bonus power budget for two reasons. First, after a period of time, the UPS requires recharge to retain its full capacity. In this case, the PDS will assign additional generation in the next period based on the required charging power. Second, PDS features an adaptive load performance scaling scheme which intelligently leverages DG energy and UPS energy for boosting the load performance (see Section 4.3 for details).

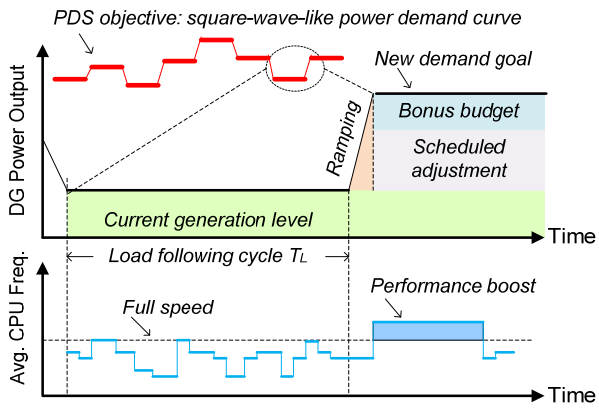


Figure 9: Power demand shaping

Definition:
T_L: Load following interval
L: Power demand of current load
P: Mean demand in the last control interval

```
for Each load tuning timestamp tick
    if tick % TL == 0          // update power budget
        Budget ← max(P, L);
    else   // maintain constant power demand
        L ← load power measurement
        if load demand changes
            ΔP_LF ← (Budget – L)
            Adjust load performance based on ΔP_LF
        end if
    end if
end for
```

Figure 10: Power demand shaping algorithm

In Figure 10 we show the pseudo code of power demand shaping. The controller monitors the data center load every tick (i.e., every second in our design) and adjusts node frequency based on the discrepancy between the budget and the original demand. At the end of each cycle T_L, the controller adjusts the budget for a new demand goal.

4.3 Adaptive Load Performance Scaling

Strict power demand shaping facilitates load following but comes with performance degradation. To this end, we propose adaptive load performance scaling that helps PDS to improve job performance while maintaining high load following effectiveness. The key idea is to use a relaxed power demand shaping policy that allows workload to explore additional energy and opportunities.

DGR Boost: The first optimization scheme helps to fine-tune the distributed generation level. Since we have no oracle knowledge of the incoming task, adjusting DG generation based on historical average unavoidably incurs prediction error. All the active jobs under this circumstance suffer low processor speed and the negative impact can last as long as the load tuning interval. To solve this problem, we propose *DGR Boost* optimization (Figure 11) which leverages the build-in over-clocking capabilities of modern processor nodes to minimize the job delay.

To enable *DGR Boost*, the PDS controller keeps the timing statistics of each active job during runtime. In any given control period, the associated computing node may spend different time t_i on different frequency F_i^{PDS} due to the time-varying demand-supply mismatch. Assume F is the full frequency without performance scaling; we define a job's normalized progress as:

$$NP_j = \frac{T}{T^{PDS}} = \frac{1}{\sum t_i}\left(\sum_i \frac{\mu F_i^{PDS} + (1-\mu)F}{F} t_i\right) \qquad (4.1)$$

In equation 4.1, T^{PDS} is the time the task spends under power demand shaping and T is the time the program would spend if no performance scaling were applied. Since frequency scaling only changes the CPU time, we use mean CPU utilization μ to estimate the proportion of runtime that is affected by power demand shaping. If *NP* always equals to 1 then PDS is equivalent to a normal job dispatcher that has no performance scaling.

In *DGR Boost* scheme, we use *NP* to evaluate the degree of job slowdown. System throughput is then defined as the mean *NP* across all the active jobs.

$$STP = \frac{1}{J}\sum_{k=1}^{J} NP_k \qquad (4.2)$$

At each end of the load following cycle, if *STP* is lower than a preset goal, the PDS controller will assign bonus power budget to the load. Meanwhile, to actually leverage this power bonus, the controller further enables CPU frequency boost mode on each node. Such frequency boost mode is well supported in the AMD Turbo Core [25] and the Intel Turbo Boost Technology [26]. By occasionally increase the CPU speed, we can catch up important deadline and avoid performance degradation incurred in strict demand shaping.

UPS Boost: This scheme fine-tunes intra-cycle load power. The main idea is that for short jobs that cannot gain the benefits of *DGR Boost*, we can leverage the UPS stored energy to avoid significant performance degradation.

In typical HPC data centers, users are required to submit their job runtime estimations to enable backfilling, which can help maximize cluster utilization. In this study we use job runtime to sort out short jobs. The short jobs here are defined as tasks that will finish before current load following cycle ends. In Figure 12, *UPS Boost* scheme will first find out all the short jobs that have low normalized progress. These jobs are very likely to miss their deadline without additional performance boost. Afterwards, the controller will check available UPS stored energy and use bin-packing method to power a group of selected short jobs.

Since *UPS Boost* can improve job performance based on accurate runtime estimates, it encourages users to not over-estimate their job runtime. One can use a pricing model [8] to further improve runtime estimation accuracy and thereby increase the effectiveness of *UPS Boost*.

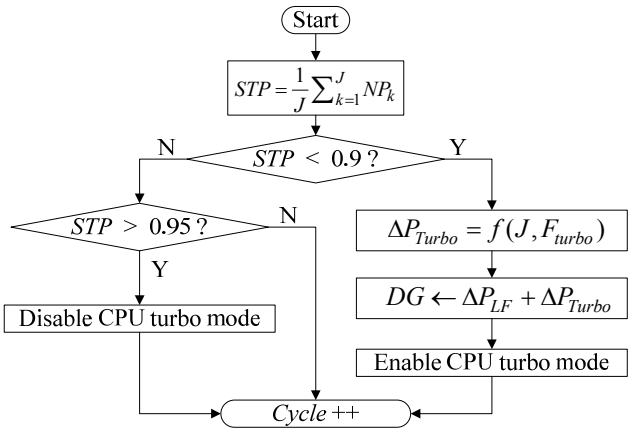


Figure 11: Flowchart of the *DGR Boost* scheme

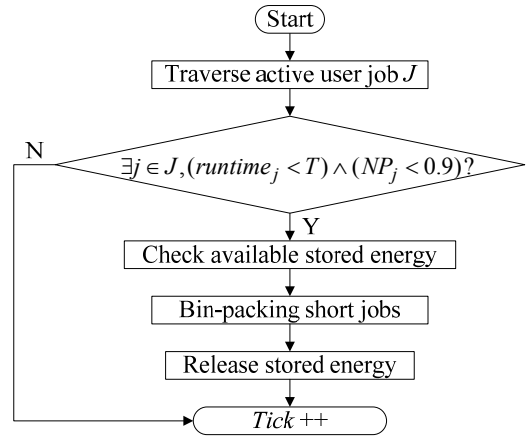


Figure 12: Flowchart of the *UPS Boost* scheme

5. Experimental Methodologies

We develop a HPC data center simulation infrastructure that takes real-world workload traces as input. Our discrete-event simulator puts each user's request into a queue and waits to grant allocation if computing nodes are available. Each job request in the trace has exclusive access to its granted nodes for a bounded duration. Such trace-driven discrete-event simulation framework has been adopted by several prior studies on investigating data center behaviors and facility-level design effectiveness [6, 27, 28].

We model an IBM System x3650 M2 (2.93G Intel Xeon X5570 processor) high-performance server which supports Intel Turbo Boost technology. While the number of processor performance states (P-states) is processor specific, we assume 12 different P-states as indicated in [29]. The minimum frequency is 1.6GHz and the normal frequency is 2.9GHz. In Turbo Boost mode, the processor could temporarily increase the CPU frequency by up to 14%. We only increase the frequency moderately (i.e., 10%) when turbo boost is enabled.

Our power model uses CPU utilization as the main signal of machine-level activity. There has been prior work showing that CPU utilization traces (sampling periods range from tens of seconds to minutes) can provide fairly accurate server-level power prediction [30]. According to the published SPEC power data, the modeled system consumes 244 Watts at 100% utilization and 76.3 Watts when idle [31]. Note that our estimates of benefits from power demand shaping are conservative; in the real world, data center workload could cause much more brief power spikes than what the simulator can capture [32], causing even worse performance degradation on non-PDS based design.

We simulate the power behavior of distributed generation system on per-second time scale which is in tune with our data center simulator. The distributed generator adjusts its output on a coarse-grained interval (10-minute as default value) and batteries respond to the fluctuating computing load. We adopt microsource optimization model HOMER in our simulator [33]. Developed by the National Renewable Energy Laboratory, HOMER simulates both conventional and renewable energy technologies and evaluates the economic and technical feasibility of a large number of power system options.

Table 2 shows the performance parameters we used in our evaluation. All the parameter values are carefully selected base on manufacturer specifications, government publications and industry datasheet. For example, we use published emission factors to calculate carbon emissions [34]. We assume a 1-minute constant duration for the distributed generation to ramp up when perform load following. The battery cycle life is set to be 10,000 times and we calculate the capacity degradation of battery cell based on its discharging profile which includes detailed information of each discharging event.

Inputs	Typical Value	Value Used
Battery life cycle	5,000 ~ 20,000 times	10,000 times
Ramping time	10 sec ~ 5 min	1 min
Rated DOD	0.8	0.8
Battery Efficiency	75% ~ 85%	80%
Battery cost	$1 ~ $3 per Ah	$2 per Ah
Emission factors of grid	0.6kg ~ 1kg CO_2/KWh	0.9 kg/KWh

Table 2: DG parameters used in the evaluation [16-20]

We use a real-world workload trace from a well-established online repository [35]. These workload logs are collected from large production systems around the world and have been scrubbed by the administrator to remove anomalous data that could skew the performance evaluation on different scheduling schemes [36]. We use five key task parameters of each trace file: job arrival time, job start time, job completion time, requested duration, and job size in number of granted computing nodes. As shown in Table 3, we select six 1-week workload traces that have different mean utilization level and mean job runtime. Our simulator uses batch scheduling, the standard method for sharing computing resources among multiple users. The job scheduling policy is first come first serve (FCFS). We examine the data center load and distributed generation budget at each fine-grained simulated timestamp.

Trace	Description	Avg. Load	Mean Inter-arrival Time	Avg. Job Run Time	
LANL	Los Alamos Lab's 1024-node connection machine	56%	4.9 min	31.6 min	Short
SDSC	San Diego Supercomputer Center's Blue Horizon HPC	60%	3.7 min	33.0 min	
Atlas	Lawrence Livermore Lab's 9216-CPU capability cluster	46%	10.6 min	36.8 min	
Seth	A 120-node European production system	85%	20.5 min	6.2 h	Long
MCNG	A 14-cluster data center, 806 processors in total	89%	2.2 min	11.1 h	
RICC	A massively parallel cluster with 8000+ processors	52%	0.9 min	16.6 h	

Table 3: The evaluated data center workload traces [35]

Scheme	Description
Oracle	Ideal power provisioning with no performance overhead
LF	Existing load following based design
ST	Existing supply tracking based design
PDS-s	PDS without optimization (i.e., uses strict power budget)
PDS-r	PDS + adaptive load tuning (*DGR Boost* & *UPS Boost*)

Table 4: The evaluated power management schemes

6. Results

In this section we discuss the benefits of applying power demand shaping to distributed generation powered data centers. In Table 4, *Oracle* is an ideal design that has a priori knowledge of load patterns and could always meet the fluctuating data center power demand with no performance overhead. It represents the optimal energy balance scenario that one can achieve with renewable energy resources. Different from *Oracle*, *LF* is a conventional load following based scheme that has heavy reliance on energy storage. *ST* represents existing power supply driven design that aims at managing the computational workload to match the renewable energy supply [4, 6, 8]. *PDS-s* is our proposed power demand shaping mechanism without optimization. *PDS-r* is the relaxed power demand shaping that features adaptive workload-aware performance scaling.

	LAN	SDSC	Atlas	HPC	MCNG	RICC
200 W/m^2	2.92	1.60	1.64	7.25	1.31	1.24
400 W/m^2	2.10	1.51	1.66	9.27	1.12	1.29
600 W/m^2	2.89	1.41	1.37	5.91	1.08	1.10
800 W/m^2	5.27	1.34	1.30	7.00	1.06	1.08

Table 5: Mean job turnaround time of *ST* under different renewable generation levels (normalized to *Oracle*)

6.1 Performance

Performance is one of the key driven forces of our power demand shaping technique. Although the then-novel concept of tracking renewable power budget has shown great success on reducing IT carbon footprint, it cannot ensure performance and sustainability simultaneously.

Table 5 evaluates the performance degradation of supply tracking based design under different wind energy intensity. We have scaled the wind power traces so that the average wind power budget equals the average data center power demand. When renewable energy generation drops or intermittently unavailable, we scale down the node frequency and defers job execution if it is necessary.

The results in Table 5 show that data centers designed to track the variable power supply may incur up to 8X job runtime increase. The geometric mean value of turnaround time increase is 59%. It also shows that workload performance heavily depends on renewable power variation behavior and data center load pattern. We have observed that occasionally the renewable energy variation pattern is totally uncorrelated with the load power demand. There is no guaranteed performance lower bound since both the user behavior and environment conditions are stochastic.

In contrast to supply tracking based design, load following fundamentally changes this situation. Figure 13 shows the mean job turnaround time under varying load following interval. All the results are normalized to *Oracle*, which has no performance scaling and maintains full-speed server operation with sufficient power budget. The average job turnaround time increase of *PDS-s* is 8.0%, which outperforms supply tracking based design most of the time. Our optimized scheme, *PDS-r*, shows 1.2% mean performance degradation compared to *Oracle*. This means that power demand shaping with adaptive load tuning (i.e., *DGR Boost* and *UPS Boost*) could yield a 37% improvement over existing supply-tracking based design.

In addition to the mean performance statistics, Figure 14 further investigates the worst-case scenario in terms of the mean job turnaround time of the worst 5% jobs. We observe that *PDS-s* and *PDS-r* shows 47% and 9.6% performance degradation, respectively. Therefore, *PDS-r* shows better performance guarantee in terms of the worst-case performance degradation.

The state-of-the-art green energy-aware scheduling schemes normally rely on utility grid as backup [4, 6, 8]. To achieve the same performance as *PDS-r*, these designs have to increase their reliance on the utility, as shown in Figure 15. The percentage of time that utility power is used is between 30% and 70%. The geometric mean value across all workloads and renewable generation levels is 51%. Heavy reliance on utility means increased carbon footprint, which goes against the original intention of sustainability.

6.2 Battery Lifetime

The reason we do not use load following directly is that existing load following scheme requires substantial support from batteries. Aggressive charging/discharging event result from IT load fluctuation shortens the lifetime of these expensive devices and depletes the stored energy quickly. Power demand shaping is advantageous because it requires

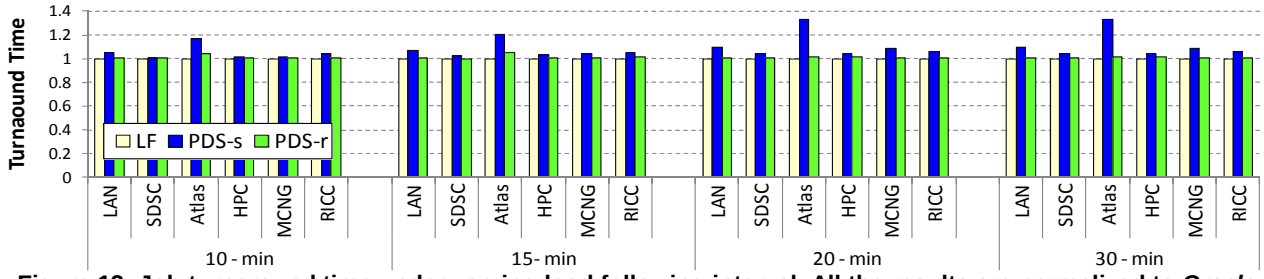
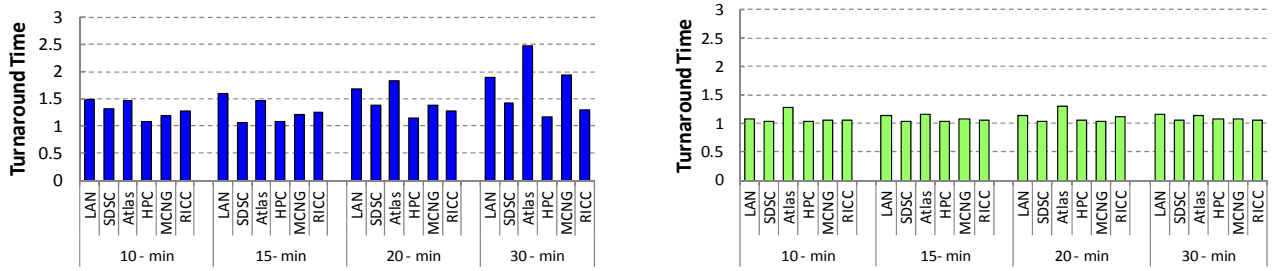


Figure 13: Job turnaround time under varying load following interval. All the results are normalized to *Oracle*



(a) PDS-s **(b) PDS-r**

Figure 14: Mean turnaround time of the worst 5% jobs. All the results are normalized to *Oracle*

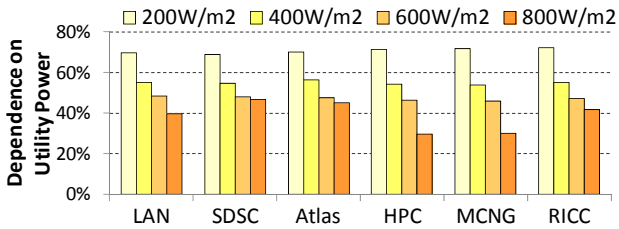


Figure 15: To compete with *PDS-r*, even state-of-the-art designs have to heavily rely on the utility power

much fewer regulation services and allows one to better explore UPS-based energy balance management. In the following discussion, we assume the UPS experiences 4 discharge cycles per day on average due to normal data center operation and maintenance needs. We evaluate the UPS lifetime based on the additional effective discharge cycles that result from load following and PDS.

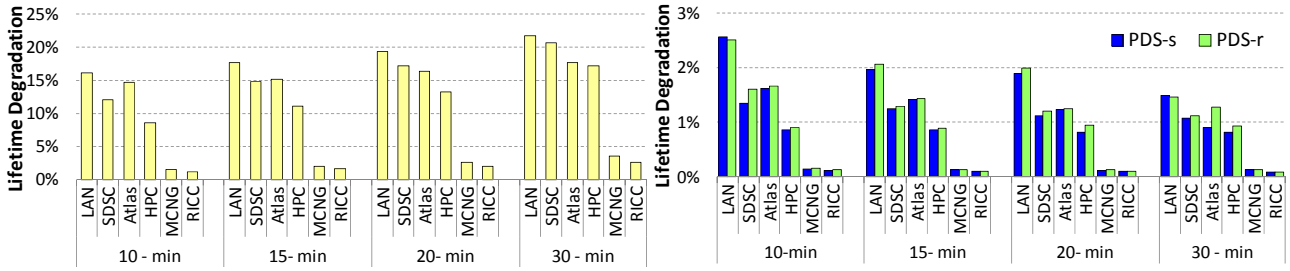
In Figure 16-(a) we evaluate the impact of load following on distributed UPS systems. The lifetime degradation due to load following varies between 3% and 21%, depending on the data center load behavior. When we increase the interval of load following cycles, the battery wear out is intensified. This is because too much of the load power fluctuation will show up as intra-cycle regulation if the time chosen for load following is too long.

Compared to conventional load following, PDS results in much lower UPS stress. In Figure 16-(b), both *PDS-s* and *PDS-r* shows less than 3% battery lifetime degradation. Although *PDS-r* explores additional stored energy to perform performance boost, its impact on UPS system is very similar to that of *PDS-s*. Note that for power demand shaping mechanism, the battery wear out problem is alleviated when we increase the interval of load following cycles. This trend is different from what we observed in *LF*. The reason is that batteries are mainly used to provide power shortfall during generator's ramp-up period in PDS based system. When we increase the load following intervals, we also lower the frequency of generator ramp-up, the dominant factor of battery discharging.

Another advantage of power demand shaping is that it allows one to use conventional centralized UPS system to assist power demand shaping. The drawbacks of centralized UPS system is that it cannot provide a fraction amount of energy – all the data center load has to be switched between the main supply and UPS. Consequently, during each UPS ride-through, the battery cells have to experience much higher discharge current. The immediate result of this is significantly decreased UPS lifetime. As shown in Figure 17, *LF* in this case results in 35% lifetime degradation on average. The lifetime degradation of *PDS-s* and *PDS-r* in this case is 6% and 7%, respectively. Therefore, we believe power demand shaping still maintains acceptable overhead and can be applied to centralized UPS systems.

6.3 Environmental Impact

We evaluate the environmental impact of DG powered data centers in terms of the annual emission savings. We mainly consider carbon dioxide which is the most important gas within the context of greenhouse gas emissions. Since the distributed generation system support a variety of fuels, the carbon footprint may vary. The results shown in Figure 18 are based on a 10MW data center. Low-carbon fossil fuels such as natural gas and diesel, although not 100% sustainable, could still reduce 38 and 12 metric tons of CO_2 per year. If we use bio-fuel (as well as hydrogen) based distributed generation, we could achieve 100 metric tons per year on average. Note that with our proposed schemes, neither of these emission savings is at the cost of decreased data center performance or availability.



(a) LF (b) PDS

Figure 16: Battery lifetime degradation under various power management schemes (with distributed UPS)

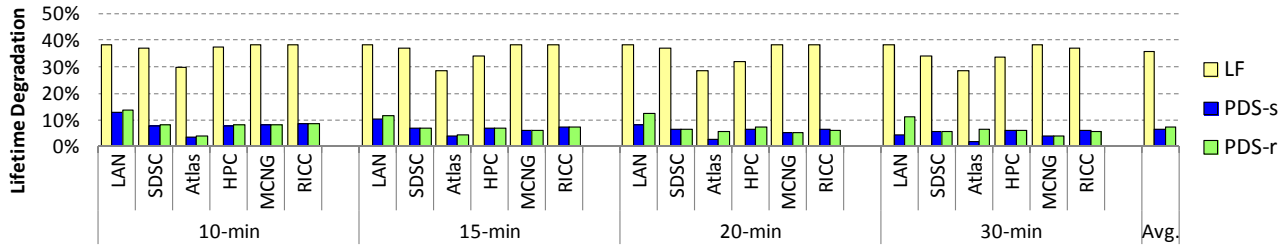


Figure 17: Battery lifetime degradation under various power management schemes (with centralized UPS)

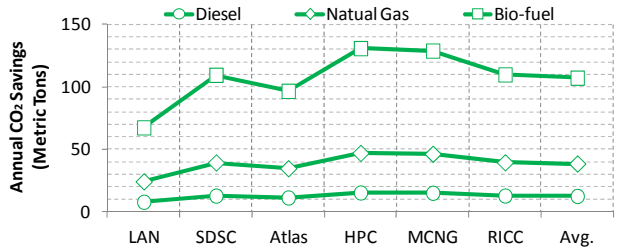


Figure 18: Annual greenhouse gas emission savings

7. Related Work

While energy-/power- aware architecture has become a research focus since long time ago, designing green data centers (especially renewable energy powered computing systems) gained its popularity only recently. Here we highlight a number of representative works that strive to improve IT sustainability in different aspects.

Objectives: The renewable energy driven computing system design is reverent to managing power variability problems in supply/load and the associated power/energy costs. For example, Li et al. [5] have investigated the optimal solar power allocation on multi-core system for maximally harvesting the solar energy and improving the overall throughput. Sharma et al. [4] have explored the performance tradeoffs between cache's hit rate and access fairness with intermittent power constraints. Goiri et al. [8, 9] focused their attention on both job deadline and power cost in a renewable energy powered data-processing cluster. Recent studies have also explored the load balancing overhead between renewable energy powered clusters and utility grid powered clusters [6, 37]. The main concern in these studies is that renewable energy power budget varies with time and the computing load must be adaptable to match the supply.

Cost-effectiveness is also a key design objective. For example, Le et.al [38] propose algorithms that minimizes fossil fuel-based energy consumptions; Liu et.al [39] discuss load balancing schemes on distributed systems. These proposals manly rely on power price arbitrage to increase the profit of data center operation. More recent work on carbon-aware energy capacity planning for data centers looked at both on-site renewable energy generations and off-site green energy purchases [40]. This work adds further motivation for us to consider distributed generation in data center design.

Mechanisms: While designing a green computing platform can be a complex undertaking, there are many techniques and mechanisms available to support the transition. Some examples include: power/cost capping [38], dynamic voltage/frequency scaling of processors [5], fast power state switching (a.k.a. power cycling) [4], virtual machine adaptation [6, 41, 42], and dynamic job dispatching [8, 9, 43], etc. At the facility level, it is often beneficial to design flexible and adaptable power management schemes to handle the power variability in supply and demand side [44].

There have been prior studies discussing the role of hardware and infrastructure in renewable energy powered systems. For example, one can leverage grid-tie inverter to provide utility backup [45] or use energy storage devices [46, 47, 48] to compensate for the intermittent renewable generation shortfall. In this paper we focus our attention on minimizing data center's reliance on the utility grid and energy storage for both sustainability and cost-effectiveness considerations. Recent proposal on data center peak power shaving leverages distributed UPS systems to improve power capping efficiency [23]. This design provides us valuable guidelines for exploring UPS-based energy balance management.

8. Conclusions

Distributed generation (DG) systems are gaining popularity and their installed capacity is projected to grow at a faster pace. We expect this trend to continue, as the energy crisis and environmental problem become increasingly crucial to our planet.

In this paper we investigate how the incoming smart grid incentive would impact the design and optimization of data centers. We propose distributed generation powered data center that leverages the load following capability of onsite renewable generation to achieve low carbon footprint without compromising performance. We have developed novel power demand shaping technique (PDS) to improve the load following efficiency in data centers while boosting the workload performance with two adaptive load tuning schemes: *DGR Boost* and *UPS Boost*. Overall, PDS could achieve 98.8% performance of an ideal oracle design and outperform existing supply-tracking based approach by 37%. In addition, our design reduces electrical stress on batteries and could improve battery lifetime by up to 26%. The unique feature of load following based design makes this work a step forward toward the goal of incorporating clean energy resources into IT systems. We expect that our work could provide valuable guidelines for data center designers in the green computing and smart generation era.

Acknowledgements

This work is supported in part by NSF grants 1117261, 0845721(CAREER), by Microsoft Research Safe and Scalable Multi-core Computing Award, and by NSFC grant 61128004. Chao Li is also supported by a University of Florida Graduate Fellowship.

References

[1] J. Koomey, Growth in data center electricity use 2005 to 2010, *Analytics Press*, 2011
[2] G. Boccaletti, M. Löffler, and J. Oppenheim, How IT can cut carbon emissions, *McKensey Quarterly*, 2008
[3] M. Arlitt et.al., Towards the design and operation of net-zero energy data centers, *IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electrical Systems*, 2011
[4] N. Sharma, S. Barker, D. Irwin, and P. Shenoy, Blink: Managing server clusters on intermittent power, *ASPLOS,* 2011
[5] C. Li, W. Zhang, C. Cho, and T. Li, SolarCore: Solar energy driven multi-core architecture power management, *HPCA,* 2011
[6] C. Li, A. Qouneh, and T. Li, iSwitch: Coordinating and optimizing renewable energy powered server clusters, *ISCA*, 2012
[7] N Deng, C Stewart, J Kelley, D Gmach, and M Arlitt, Adaptive green hosting, *ICAC*, 2012
[8] Í. Goiri, R. Beauchea, K. Le, T. Nguyen, M. Haque, J. Guitart, J. Torres, and R. Bianchini, GreenSlot: scheduling energy consumption in green datacenters, *Supercomputing*, 2011
[9] Í. Goiri, K. Le, T. Nguyen, J. Guitart, J. Torres, and R. Bianchini, GreenHadoop: Leveraging green energy in data-processing frameworks, *Eurosys*, 2012
[10] R. Lasseter and P. Piagi, Microgrid: A Conceptual Solution, *IEEE Annual Power Electronics Specialists Conference*, 2004
[11] The role of distributed generation and combined heat and power (CHP) systems in data centers, *Technical Report*, EPA, 2007
[12] B. Kirby and E. Hirst, Customer-specific metrics for the regulation and load-following ancillary services, *Technical Report*, ORNL, 2000
[13] http://www.apple.com/environment/renewable-energy/
[14] Microsoft looking to test grid-independent data center, http://www.datacenterdynamics.com
[15] eBay plans data center that will run on alternative energy fuel cells,http://www.nytimes.com/2012/06/21/technology/
[16] L. Schwartz, Distributed generation in Oregon: overview, regulatory barriers and recommendations, *Technical Report*, Oregon Public Utility Commission, 2005
[17] Fuel cell technologies program multi-year research, development and demonstration plan, *Technical Report*, U.S. DOE, 2012
[18] H. Zareipour, K. Bhattacharya, and C. Canizares, Distributed generation: current status and challenges, *The 36th Annual North American Power Symposium*, 2004
[19] Battery technology for data centers and network rooms: VRLA reliability and safety, *APC Write Paper,* 2003
[20] S. Drouilhet and B. Johnson, A battery life prediction method for hybrid power applications, *Technical Report*, NREL, 1997
[21] S. Chowdhury and P. Crossley, Microgrid and active distribution networks, *The Institute of Engineering and Technology*, 2009
[22] P. Piagi and R. Lasseter, Autonomous control of microgrids, *IEEE Power Engineering Society General Meeting*, 2006
[23] V. Kontorinis, L. Zhang, B. Aksanli, J. Sampson, H. Homayoun, E. Pettis, T. Rosing and D. Tullsen, Managing distributed UPS energy for effective power capping in data centers, *ISCA*, 2012
[24] A. Gandhi, M. Harchol, R. Das, and C. Lefurgy, Optimal power allocation in server farms, *SIGMETRICS,* 2011
[25] The new AMD Opteron™ processor core technology, AMD
[26] http://www.intel.com/content/www/us/en/architecture-and-technology/turbo-boost/turbo-boost-technology.html
[27] S. Pelley, D. Meisner, P. Zandevakili, T. Wenisch and J. Underwood, Power routing: Dynamic power provisioning in the data center, *ASPLOS,* 2010
[28] F. Ahmad and T. Vijaykumar, Joint optimization of idle and cooling power in data centers while maintaining response time, *ASPLOS,* 2010
[29] Host power management in VMware vSphere 5, VMware, 2010
[30] P. Ranganathan, P. Leech, D. Irwin, and J. Chase, Ensemble-level power management for dense blade servers, *ISCA*, 2006
[31] SPECpower_ssj2008, http://www.spec.org/power_ssj2008/
[32] D. Meisner, and T. Wenisch, Peak power modeling for data center servers with switched-mode power supplies, *ISLPED*, 2010
[33] Getting started guide for HOMER version 2.1, National Renewable Energy Laboratory, 2005
[34] Unit conversions, emissions factors, and other reference data, *Technical Report of EPA*, 2004
[35] http://www.cs.huji.ac.il/labs/parallel/workload/logs.html
[36] D. Tsafrir and D. Feitelson Instability in parallel job scheduling simulation: the role of workload flurries, *IPDPS,* 2006
[37] C. Li, A. Qouneh, and T. Li, Characterizing and analyzing renewable energy driven data centers, *SIGMETRICS*, 2011
[38] K. Le, O. Bilgir, R. Bianchini, M. Martonosi, and T. D. Nguyen, Capping the brown energy consumption of Internet services at low cost, *IGCC*, 2010
[39] Z. Liu, M. Lin, A. Wierman, S. Low and L. Andrew, Greening geographical load balancing, *SIGMETRICS,* 2011
[40] C. Ren, D. Wang, B. Urgaonkar, and A. Sivasubramaniam, Carbon-aware energy capacity planning for datacenters, *MASCOTS*, 2012
[41] A. Kansal, F. Zhao, J. Liu N. Kothari and Arka A. Bhattacharya, Virtual machine power metering and provisioning, *SOCC,* 2010
[42] P. Lama and X. Zhou, NINEPIN: Non-invasive and energy efficient performance isolation in virtualized servers, *DSN*, 2012
[43] Y. Zhang, Y. Wang, X. Wang, GreenWare: Greening cloud-scale data centers to maximize the use of renewable energy, *Middleware*, 2011
[44] P. Banerjee, C. Patel, C. Bash, and P. Ranganathan, Sustainable data centers: Enabled by supply and demand side management, *DAC*, 2009
[45] N. Deng, and C. Stewart, Concentrating renewable energy in grid-tied datacenters, *ISSST*, 2011
[46] S Govindan, A. Sivasubramaniam, and B. Urgaonkar, Benefits and limitations of tapping into stored energy for datacenters, *ISCA*, 2011
[47] R. Urgaonkar, B. Urgaonkar, M. Neely, and A. Sivasubramaniam. Optimal power cost management using stored energy in data centers, *SIGMETRICS*, 2011
[48] S. Govindan, D. Wang, A. Sivasubramaniam, and B. Urgaonkar, Leveraging stored energy for handling power emergencies in aggressively provisioned datacenters, *ASPLOS*, 2012