UPS-Aware Workload Placement in Enterprise Data Centers

Quan Zhang and Weisong Shi
Department of Computer Science
Wayne State University
Emails: {quan.zhang, weisong}@wayne.edu

Abstract

Energy efficiency has become a very important concern for enterprise data centers due to the significant cost on electricity. To improve the energy efficiency of data centers, researchers and practitioners have proposed enormous work to reduce the energy consumption of data centers. Most of previous work focus on reducing the energy consumption of IT equipment; however, the power losses caused by uninterruptible power supply (UPS) is not considered, which could account for 15% of the total energy cost of a data center. Recent trend shows that rack level UPSes are getting popular. In this work, we focus on power minimization for both IT equipment and power losses of rack level UPS in an enterprise data center. We propose a rack level power model that builds a direct relationship between IT workload and its power dissipation. Based on this rack level power model, we formulate a mathematical formula for the workload placement optimization problem. The experimental results show that the rack level power model precisely matches the measured power, and the error rate is within ±2.5%. The simulation also indicates up to 5.2% power saving compared to uniform workload allocation, which means $1.425 million savings for a 76 megawatts data center with PUE 1.7.

1 Introduction

Energy efficiency has become a very important concern for data centers due to the significant cost on electricity. Data centers are consuming a staggering amount of energy every year making energy costs one major contributor of their total cost of ownership (TCO). According to NRDC, the data centers in US alone consume 91 billion kWh of electricity in 2013, and it will consume 140 billion kWh annually by 2020 [1]. To improve the energy efficiency of data centers, researchers and practitioners have proposed many techniques to reduce the power consumption of IT equipment. The most popular approach employs dynamic voltage/frequency scaling (DVFS) to reduce power dissipation of the CPU and memory subsystem [2,3]. Server consolidation is another approach for power reduction of IT equipment, which assigns tasks to fewer servers and shuts down idle servers [4,5]. Load balancing refers to allocating total workload among servers evenly in order to balance the per server workload [6,7]. In order to maximize energy efficiency, understanding the relationship between resource usage and system-level power consumption is mandatory for data center operators. The power models for specific components of a computer, such as CPU, memory, and disk subsystems, are proposed based on the hardware resource utilization [8,9,10]. Based on the subsystems power modeling, the system level power modeling have also been explored for non-virtualized and virtualized environment [11,12,13]. Although previous approaches have managed to reduce the power consumption of IT equipment that contributes 30% of total data center energy cost, the power losses of an uninterruptible power supply (UPS), which accounts for another 15% of total energy cost [14], has not been explored.

UPS Background Figure 1 shows a simplified power flow in a typical data center. At the highest layer, the utility power and the backup power (e.g., diesel generator) is passed through UPS units via automatic transform switch (ATS). Power then goes through Power Distribution Units (PDUs) to different racks, which then gets distributed through power strips to individual servers which all have their own power supplies. To ensure the availability of data centers, a redundant UPS deployment is commonly used for both centralized and distributed UPS topologies since UPS is a single point of failure. UPS configurations are often described by nomenclatures using the letter N in a calculation stream, where an N system is a system comprised of a single UPS module, or a paralleled set of modules whose capacity is matched to the critical load projection. Two popular redundant configurations are N+1 redundant (a.k.a. parallel redundant) and 2N redundant (a.k.a. system plus system redundant).
A parallel redundant configuration consists of paralleling multiple, same size UPS modules, and the system is N+1 redundant if the spare amount of power is at least equal to the capacity of critical load. System plus system redundant is the most reliable and expensive design in the industry, and a 2N system can tolerate every conceivable single point of failure.

Based on the UPS redundant configuration, the power loss behaviors are completely opposite: for N+1 UPS configuration, the power losses decrease when the IT power load increases; while for 2N UPS configuration, the power losses increase along with IT power load increasing. From the view of rack level UPS, using fewer servers running at the full speed, or using more servers at the lower speed with uniform workload distribution is not always power saving since lowering UPS output load leads to lower conversion efficiency. This observation motivates this work as described in the following section. Enterprise data centers usually run fewer applications, and sometimes just one that is spread across entire data center, such as Web 2.0 and Software as a Service (SaaS) deployed at Google’s data centers. The workload of enterprise data center scale is relatively high, and virtualization is often not required so this can mean a single data center could have tens of thousands of physical servers. For such enterprise data centers, the workload placement is more important since millions of users’ requests need to be separated. Thus, in this paper we focus on the workload placement of enterprise data centers.

In this paper, we present an optimization problem that minimizes the energy cost of IT equipment and UPS(es). We first build a rack level power model mapping the workload directly to its power dissipation. Second, we formulate a mathematical problem that chooses an optimal workload allocation for minimizing the power consumption of IT equipment and wasted power of UPS(es). Finally, we analyze the possible electricity cost saving using a total cost of ownership (TCO) model. The experimental results show that the rack level power model precisely matches the measured power, and the error rate is within ±2.5%. We also simulate a data center hosting 50 racks (1000 servers) with 10 applications. The simulation indicates up to 5.2% power saving compared to uniform workload allocation, which means $1.425 millions/yr energy cost savings for a 76 megawatts (MW) data center with power usage effectiveness (PUE) 1.7.

The remainder of this paper is organized as follows. In Section 2, we analyze the power losses of different UPS topologies. Section 3 presents the rack level power model and the optimization problem. Section 4 shows the experimental and simulation results, and total energy cost saving analysis. Finally, Section 5 concludes.

2 UPS Power Losses Analysis

Within an UPS, power is lost when it is transformed between AC and DC. For a double conversion UPS, there are two places where power is lost. The first is when power is transformed from AC to DC to be stored in batteries, and the second is when power is transformed from DC to AC to be delivered to racks and servers. For UPS topology, a centralized topology usually deploys UPS at the facility level, while a distributed topology deploys UPS at rack level or server level [15]. The choice of appropriate redundant configuration is based on how often failures happen in a data center.

In this paper, we focus on rack level distributed UPS topology and its power losses. The power loss of UPS is determined by loaded capacity, UPS efficiency curve, and UPS redundant configuration. The loaded capacity depends on the IT workload. For N+1 UPS configuration, the loaded capacity varies from 0% to 100%, while for 2N configuration, the maximum loaded capacity is only 50% since the total power load is evenly allocated to two UPSes. UPS efficiency curve depends on the technology that manufacture uses. Figure 2 shows an instance of UPS efficiency curve, which is produced using the data collected from our UPS system. Usually, lower UPS effi-
Figure 3: Power losses of a single rack with N+1 and 2N configurations.

Efficiency leads to higher power losses within UPS. In data centers, the IT equipment load decides the loaded capacity and consequently the power losses of UPSes. To get the relationship between IT equipment power and power losses of UPS, we use an UPS with a rating power of 8 kilowatts to supply the power of a rack with 20 servers. All servers have the same measured peak power of 350 watts. The idle power of 20 servers is 4,366 watts.

Given the UPS efficiency curve showed in Figure 2, we first use a natural logarithm function to fit the curve and calculate the UPS power losses using Mathematica. Figure 3 shows the power losses of UPS(es) with N+1 and 2N configurations. For N+1 configuration, the power losses increase when the loaded capacity is less than 50% and then decrease when the loaded capacity is higher than 50%. For 2N configuration, the power losses continuously increase along with the increasing of loaded capacity. To show that the optimal workload distributions for different UPS configurations are different, we use two racks as described above with 20 servers fully loaded and the other 20 servers fully idle. Two workload distributions are used: (1) 20 fully loaded servers are on the same rack, and all the idle servers are on the other rack; and (2) 12 fully loaded servers and 8 idle servers are on one rack, and the remaining servers are on the other rack. Table 1 shows the UPS power losses of different UPS configurations and workload distributions. For N+1 configuration, the workload distribution type 1 has lower power losses, while for 2N configuration, the workload distribution type 2 has lower power losses. Based on this observation, we formulate an optimization problem to minimize the total power of IT equipment and UPS power losses.

### 3 Energy Efficient Workload Placement

In this section, we first present the rack level power model which directly maps the entire rack workload to its power dissipation. Next, we give a mathematical formulation for workload placement problem in data centers, which minimizes power consumption of both IT equipment and power losses of UPS(es).

#### 3.1 Rack Level Power Modeling

Traditional power models, which usually leverage the performance counter and utilization information, focus on server level power estimation. In this paper, we propose a rack level power model that directly uses the workload information, such as throughput and instructions per second (IPC), as the inputs of the power model. In this paper, we build the power model for an enterprise data center where servers are not virtualized and each server only hosts one application. We also assume the CPU is running at a fixed speed without dynamic tuning.

The power model used for a rack can be expressed as in (1)

$$ P_i(w) = P_i^{IDLE} + \sum_j \alpha_j^i \times w_j^i, $$

where $P_i^{IDLE}$ is the idle power of a rack, and the summation of $\sum_j \alpha_j^i \times w_j^i$ is the total power introduced by all workloads on this rack. The $\alpha_j^i$ is the coefficient that represents the watt per performance of workload $j$ on the $i^{th}$ rack. This $\alpha_j^i$ has different units for different applications and hardware. For CPU-intensive applications, the metric of $\alpha$ could be $\text{watts/instruction}$; for memory-intensive applications, the metric could be $\text{watts/byte}$; and for web services, $\text{watts/request}$ could be a good indicator to show how efficient the system is. Workload profiling provides prior or historic knowledge that can be used to choose the $\alpha$. In this paper, we profile an application with the following steps. First, we measure the idle power of the $i^{th}$ rack as $P_i^{IDLE}$. Second, we fully load the rack and get a rack level performance upper bound for this application. Third, we gradually increase the workload making the rack run at different power levels and...
record the rack power. Finally, we calculate the average value (performance per watts) of all sample points, which is used as the \( \alpha_i \) for the application. We repeat this process for different types of applications and get their corresponding \( \alpha_i \) value for the \( i^{th} \) rack.

### 3.2 Optimization Problem Statement

We consider the optimization problem for a data center hosting multiple applications simultaneously. We assume that each server only hosts one application at any time, and we can dynamically assign the workload to a different subset of servers. In addition, we assume that one UPS is only connected to one rack and the UPS is deployed in either N+1 or 2N redundant configuration. As we mentioned above, the power losses of UPS(ES) vary a lot with IT power load vibration. Clearly, any change of the workload or workload distribution will affect the power of IT equipment and power losses.

The goal is to minimize the total power of both IT equipment power and wasted power in rack level UPS(ES) by choosing optimal workload allocation given the equality and inequality constraints of (i) performance constraint, which means the summation of all racks’ workload should be equal to the total workload submitted by all users; (ii) capacity constraint, which means the hardware resource requirement should be less than the maximum hardware capacity in each rack; and (iii) power constraint, which means the total rack power satisfies the power capping requirement specified by data center operator or hardware. Based on this description, a mathematical formulation of the optimization problem can be expressed as follows:

\[
\text{Minimize} \quad \left\{ \sum_i P_i \right\} \quad \text{(2)}
\]

\[
\sum_j w_{j}^i = W^j \quad \text{(3)}
\]

\[
\sum_j \frac{w_{j}^i}{C_i} \leq C_i \quad \text{(4)}
\]

\[
P_i^{\text{IDLE}} + \sum_j \alpha_i \times w_{j}^i \leq P_i^{\text{CAP}} \quad \text{(5)}
\]

where \( P_i \) is the power of the \( i^{th} \) rack, and the \( \eta(P_i) \) is the conversion efficiency when the UPS has the IT power load of \( P_i \). Equation (3) ensures the performance requirement is satisfied for workload \( j \). In (4), the \( C_i \) is the capacity limitation for workload \( j \) on rack \( i \), and \( C_i \) is the hardware limitation for rack \( i \). The \( P_i^{\text{CAP}} \) in (5) is the capping power for the \( i^{th} \) rack.

The function \( \eta(P_i) \) indicates the relationship between UPS output power load and its corresponding conversion efficiency. The \( \eta(P_i) \) can be expressed as following:

\[
\eta(P_i) = a \times \ln\left( \frac{P_i}{P_{\text{UPS}}} \right) + b \quad \text{(6)}
\]

where \( P_{\text{UPS}} \) is the input power of UPS, and \( a \) and \( b \) is constant to match the conversion efficiency curve for different UPSEs. In this paper, we choose \( a \) and \( b \) of 0.1279 and 0.9343, respectively.

### 4 Evaluation Results

#### 4.1 Power Model Accuracy

To verify the power model, we conduct an experiment that uses 10 servers and 2 applications. These servers include 8 INTEL CPU servers and 2 AMD CPU servers. The two applications are Y-Cruncher [16] and Yahoo Cloud Serving Benchmark (YCSB) [17]. Y-Cruncher is a CPU-intensive application, and YCSB is used to simulate the web service requests to read and write database. First, we run those two application separately to get the \( \alpha \) as described in [1]. After that, we estimate the real time power by running two applications on all ten machines at the same time, and we randomly change the application’s workload during the test. The sample frequency is 1 Hz, and it is sufficient for long execution time tasks which are in the order of hours or days. We conduct the test on INTEL and AMD machines separately and use linear regression to fit the data points. The \( P_i^{\text{IDLE}} \) of 10 machines is 2183 watts. The coefficients \( \alpha \) of Y-Cruncher are 3E-5 watts/digits/second and 4E-5 watts/digits/second for INTEL and AMD servers, correspondingly. 0.0024 watts/ops/second and 0.0039 watts/ops/second are the coefficients \( \alpha \) of YCSB on INTEL and AMD servers. The workload of Y-Cruncher and YCSB are represented in digits/second and ops/second.

Figure 4 shows the measured power by a power meter and the estimated power using our rack level power model. The results shows that the error rates are within \( \pm 2.5\% \), which in terms of watts means the power estimation error is less than 83 watts. Furthermore, our rack level power model overestimates the power consumption in 82\% time duration (247 out of total 300 sample points). For those underestimated cases, the gap between measured power and estimated power is less than 47 watts with an error rate of \( -1.4\% \). This observation is important because there is a power capping constraint in the optimization problem. High underestimation probability and error rate can lead to violation of rack level power capping requirement.
4.2 Simulation Results

Based on our rack level power model, we compare total power consumption of both IT equipment and wasted power within UPS(es) of our optimal workload allocation, with the power consumption of a baseline case that evenly allocates workload among all racks. In the simulation, we assume that each rack hosts 20 servers, and for each rack the power is supplied by one or two UPS(es) with N+1 or 2N UPS configuration, respectively. The total number of racks is 50, and there are 10 different applications running among all racks simultaneously. To solve the optimization problem, we use Mathematica to perform our simulation. The simulation stops when either the iterations exceed a predefined threshold or the results converge to a requested precision.

We load the data center with different workload levels and solve the optimization problem which minimizes the total power consumption. Figure 5 shows the power reduction of our optimal workload allocation compared to even workload allocation. For both N+1 and 2N UPS configurations, the optimal workload allocation reduces the power consumption by 1.23% to 5.2%. Both UPS configurations have similar power reduction rates, and N+1 configuration has higher power reduction than 2N configuration. However, the power reduction gap between N+1 and 2N is small for all loaded levels. Another important observation is that our optimal workload allocation achieves the highest power reduction at the data center utilization of 50%, which is also the average utilization level of most data centers [18]. The power reduction depends on the UPS efficiency curve and the coefficient $\alpha$ in Equation 1. As an extreme example, if the UPS efficiency is constant, the power consumption is the same for all data center utilization. No matter how the workload is distributed, for certain total amount of workload, constant UPS efficiency results to same power consumption (i.e. total output power of UPS(es)). Since the UPS efficiency is constant, the input power of UPS is constant and the power losses are also constant. In this case, compared to the even workload allocation, the power reduction is zero. On the other hand, the coefficient $\alpha$ decides the power increase rate for a specific application, and consequently it affects the power consumption of data center (i.e. the output power of UPS(es)). Thus, different workload allocation may consume district power given certain total amount of workload.

To get a deep understanding of how workload type affects the power reduction in our optimization problem, we change the mixing proportion of applications while keeping the total data center utilization at 50%. In this simulation, all ten applications are divided into two categories: CPU-intensive applications and web-service applications. We mix those two types of applications with different proportions and get the corresponding power reductions. Figure 6 shows the power reduction for different mix proportions of CPU-intensive applications (denoted as CPU) and web-service applications (denoted as Web) when the data center is loaded at 50%. Figure 6 indicates more CPU-intensive workload introduces higher power reduction and the maximum power reduction is achieved at a mixing proportion of (CPU:Web = 0.85:0.15). The reason for this observation is that more CPU-intensive workload contributes more power to the rack, and higher rack power increases the loaded capacity of UPS which may lead to less power losses.

4.3 Energy Cost Analysis

The power capacities of data centers vary by a lot. Based on the estimates of power demand at 2011, the power capacity of a data center could be as low as 2 MW and up to 100 MW [19]. Fifty percent of these data centers are be-
tween 20 MW to 76 MW. According to the 2014 Uptime Institute report [20], the average PUE of data centers is 1.7. For a 76 MW data center, the power consumption of IT equipment will be 44.7 MW. With the maximum power consumption reduction of 5.2% shown by our simulation, the annual energy cost of such a data center can be reduced by $1.425 millions.

5 Conclusions

In this paper, we explore how different UPS configurations affect the energy efficiency of data centers. For different UPS configurations, the power losses of UPS have different trends when the IT workload changes. Based on this observation, we build a power model that directly maps the IT workloads to its power dissipation, and conduct experiments to show that a small error rate can be achieved using our power model. To minimize the power consumption of IT equipment and power losses of UPS, we also formulate an optimal workload placement problem, which takes the energy losses of UPS into account. The simulation shows up to 5.2% power reduction compared to an even workload allocation strategy. The energy cost analysis indicates $1.425 millions in savings each year for a 76 MW data center with PUE of 1.7. For future work, we will focus on how DVFS and switching on/off servers affect the UPS efficiency, and explore how to apply those techniques with the consideration of UPS efficiency.

Acknowledgment

This work is in part supported by NSF grant CNS-1205338. We thank the great support from Computing and Information Technology Department of Wayne State University and NextEnergy for data collection and running experiments. This material is based upon work supporting while serving at the National Science Foundation.

References


