Power Grab in Aggressively Provisioned Data Centers: What is the Risk and What Can Be Done About It

Xiaofeng Hou, Luoyao Hao, Chao Li, Quan Chen, Wenli Zheng, and Minyi Guo
Department of Computer Science and Engineering, Shanghai Jiao Tong University
{xfhelen, haoluoyao}@sjtu.edu.cn, {lichao, chen-quan, zheng-wl, guo-my}@cs.sjt u.edu.cn

***Abstract*—Aggressively provisioned data centers achieve great cost savings by over-committing the very expensive power distribution infrastructure. However, existing proposals for managing load power demand in such a data center are largely utilization-driven, overlooking power-related interferences among users. An important observation is that some tasks can impact existing power budget management framework and disrupt normal operation by taking away the precious public power capacity. This vulnerability exposes data centers to a new type of risk that we call power grab, which is essentially hostile power resource competition. It could worsen the performance-utilization tradeoff in a power-constrained computing environment.**

Anticipating a growing case for power-oriented com-petition, we propose CFP, a resilient power capacity management framework for improving the fairness and service quality in scale-out data centers. Our solution features a market-based power resource allocation and billing scheme that involves users in the loop. It allows the data center to bypass the formidable task of identifying malicious users and defend against power grab with reward and punishment incentives. We build a proof-of-concept system and also evaluate our design with realistic Google cluster traces. Compared to prior arts, CFP can increase the average performance-cost ratio by 1.8X. It can boost the total throughput in an APDC by 15% under severe power contention. Our design allows scale-out data centers to safely exploit the benefits that power over-subscription may provide, with minor overhead.

Keywords*— *power oversubscription; data centers; resource contention; market mechanism; power management.

I. Introduction

Designing aggressively provisioned data center (APDC) has attracted great attention in the past several years [1-4]. With an emphasis on optimizing data center utilization, APDC over-subscribes non-IT infrastructure such as power distribution system, cooling system, energy backup, etc. Considering that the non-IT infrastructure can be expensive and difficult to expend [5, 6], APDC shows great advantages in accommodating the fast-growing scale-out workloads.

Despite many benefits, over-subscribed data centers show an obvious limitation. Different compute instances that share public power resources face unconventional interferences as power budget shrinks. If certain user or a combination of tasks consume disproportionate amount of power, the APDC might not have enough power headroom to handle the demand surge. To prevent electrical overload and a tripped circuit, it is common practice to cap the power drawn and throttle resource usage [3, 7]. Unfortunately, most power capping methods do not directly apply to individual users or virtual machines. In other words, it can cause cluster-wide performance degradation.

The above issue becomes even acute if there are dominant power consumers causing fierce power capacity competition. Particularly, we consider an adversary that manipulates its load and abuses power resources to impact the service quality of its competitors. In this study we refer to this type of malicious act as power grab (PG). It aims to originate the worst-case performance-utilization tradeoff in an APDC.

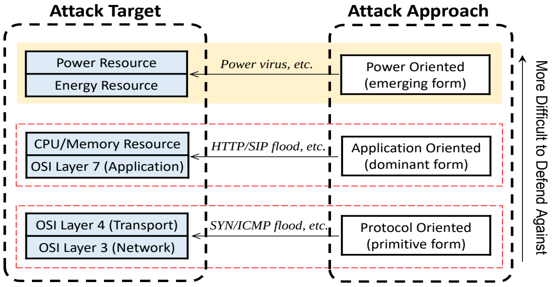


Fig.1. Resource availability threat by layer. Top-layer competitions are often harder to defend against.

From the viewpoint of resiliency, the "utilization-first" design philosophy of APDC unavoidably makes it susceptible and vulnerable to power-oriented attack. It has been shown that the proportion of unplanned data center power outage due to cyber-attack has escalated over the past six years — from just 2% in 2010 to over 20% in 2016 [8]. As shown in Figure 1, power-oriented attack would be a trending method that can interrupt data center operation. Many defense mechanisms today mainly looks at the net-working level (i.e., how IT resources are accessed) [9] and the OS level (i.e., how IT resources are utilized) [10]. Very limited work has been done at a level concerning how IT resource usage pattern may frustrate power management strategies and in turn affect IT operation.

There have been a few pioneering works that discuss power-oriented attacks such as energy abuse [11, 12] and power attack [3, 13]. Energy abuse mainly focuses on frustrating efficiency optimization efforts. It is a relatively mild attack that has limited impact on a data center. Power attack aims at causing costly down-time. It is largely an opportunistic attack that shows limited success rate. In contrast, power grab is much easier to launch and it has broad impact on workload performance and data center operators' reputation.

Power grab in power-constrained data centers presents existing power management framework with an embarrassing situation. If we ignore it and use conventional rigid power capping strategy, peak power shaving activities can at the same time cause collateral damage (unnecessary performance scaling) on normal tasks. On the other hand, tracking fine-grained load power usage and simple limiting PG user along is not appropriate. Power-hungry users can be normal users that run high-priority or time-sensitive jobs. Since it is hard to define what constitutes the attributes of "legitimate" users, detecting and capping malicious power resource consumption is almost an impossible task.

TABLE I. PEAK POWER CAPPING IN DATA CENTERS

Data Source	Peak Height	Capping Level	Capping Time
Google	~40%	Data center	1~2% runtime in total
Facebook	~30%	Switch board	10~20 min.
Microsoft	~20%	Data center	95% peaks < 4 min.

To date, most of the data center services are priced solely based on the occupancy of IT resources (i.e., CPU hour, memory, and bandwidth), but ignores the indirect competition of non-IT resources (e.g., power capacity etc.) among resource-intensive users. This policy indirectly motivates malicious users to abuse non-IT resources.

We argue that the growing workload and shrinking power budget exacerbate power resource contention in today's data center. A lack of user interaction has led to a situation where the high-level power allocation policies seem to be reasonable from a data center's perspective but it is unfair to most of its tenants/users. It impels us to think about "how data center should be designed to preserve the substantial cost benefits of aggressive power provisioning without making it vulnerable to power-oriented attacks.

We propose CFP (Charges for Power, or recursive acronym "CFP fines PG!"), a novel framework for enhancing the resiliency of today's power-constrained data centers. The key idea behind our design is to put consumers into the data center power management loop and charge them for power capacity usage. We neither blindly cap power usage nor seek effective detection method of malicious users. Instead, we focus on exposing detailed user-level power variability to the data center and creating differentiated services for data center users/tenants. Specifically, our differentiated service considers user's willingness of power scaling and leverages a bidding-based power resource auction to manage the gap between power supply and demand.

We assume all users are cost-conscious. Our APDC pricing-based power management mechanism gives rewards to users that actively participate in power budget scaling and increases the cost for potential power grab users that demand extensive amount of power. CFP is largely orthogonal to other system and architecture level power management mechanisms and can be implemented in a light-weight manner. It can thwart the resource starvation attempts of attackers while maintaining fairer resource allocation among normal users.

This paper makes the following contributions:

- We investigate power grab, a new class of threat to APDC. We discuss the source of vulnerability in detail.
- We propose a risk mitigation scheme called CFP. It alleviates power grab by encouraging open competition for priced power resources in the data center.
- We implement our design as a proof-of-concept system. We also evaluate it with extensive simulation using realistic Google cluster traces.

The rest of this paper is organized as follows. Section 2 introduces source of vulnerability. Section 3 discusses feasibility of power grab. Section 4 proposes CFP. Section 5 describes experimental methodology. Section 6 presents evaluation results. Section 7 discusses related work and Section 8 concludes this paper.

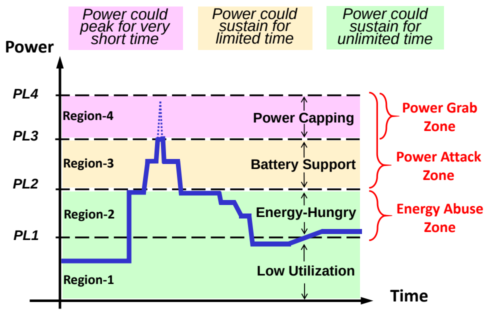


Fig.2. Power-oriented attacks from the perspective of key power management thresholds. (PL4: Max Power; PL3: Backup Limit; PL2: Normal Budget; PL1: Min Power).

II. SOURCE OF VULNERABILITIES

The rise in power grab stems from a rigid and aggressive power management framework. In this section we introduce detailed background and discuss the vulnerability of an APDC.

A. *Implicitly Defined Power Limits*

APDC leverages power budgeting (capping) strategies to improve utilization. Power capping refers to any control techniques that allow the system to stay within certain power limit. Table I shows several examples of peak power capping activities from Google [14], Facebook [15], and Microsoft [16]. Through power capping, data centers can multiplex the given power budget and enable more aggressive server over-provisioning. The insight here is the statistical effect that simultaneous peak activity across a large group of nodes is rare.

In recent years, energy storage devices (uninterruptible power supply, UPS) are also used for occasional peak power suppression [1, 2]. A group of nodes can be temporarily powered by local battery systems during peak (e.g., peak shaving) until the stored energy becomes inadequate.

Typically, current APDCs are configured with three power threshold values, as shown in Figure 2:

- *Min Power (PL1):* The average power demand in a data center. The typical value is about 30~60% of PL4 [14]. Due to the dramatically increased power capping activities, the designed power capacity rarely drops below this threshold.
- *Normal Budget (PL2):* Refers to the upper bound value of load power demand. In an APDC, PL2 is the maximum output that the power delivery system can support. It is sometimes a soft limit on power consumption since UPS batteries can be used to temporarily shave small peaks.
- *Backup Limit (PL3):* The value of PL3 may vary, depending on the reserved backup energy. PL3 is small than PL4 when the backup power system is under-provisioned for cost saving [17]. When the load power exceeds PT3, power capping must be applied. This is because even if all the UPS starts to discharge, it cannot completely shave the peak.
- *Max Power (PL4):* The aggregated server name-plate power. It is the highest power consumption that the data center can achieve theoretically. No power capping activity is required at any time if the given power budget is PL4.

TABLE II. RAW DATA FROM GOOGLE CLUSTER

Trace characteristics	Value
Time span of trace	29 days
Amount of jobs	650k
Numbers of users	925
Submitted tasks	25M
Scheduler events	143M
Resource usage records	1233M

B. *Power/Energy Oriented Attacks*

Given the above power limits, one can logically divide the power demand plot area into four regions, as shown in Figure 2. By manipulating load power demand, a sophisticated adversary can launch different types of power/energy oriented attacks.

Energy attack [11, 12], or energy abuse attack, generally falls into Region-2 which indicates high server utilization. Since resource contention is common in this region, a malicious user can easily disrupt the operation of normal tasks (e.g., force LLC miss in shared memory systems), resulting in increased runtime and energy consumption. This method can be less effective when server utilization is low (Region-1).

Power attack [3, 13], or power overload attack, aims at causing tripped circuit breaker in APDC. As shown in Figure 2, it involves two steps that spans across two zones. The attacker must first generate wide power peaks in Region 3. This will trigger peak power shaving and deplete the reserved backup energy. Afterwards, the attacker needs to stress the power system opportunistically with carefully controlled peak power shape [3]. The APDC may enforce power capping to defeat the attack once the power peaks are detected.

Different from power and energy attack, power grab mainly falls into Region 4. Compared to power attack, power grab is easier to launch. It does not emphasize overload and there are no restrictions on the shape of power peaks in the Region 4. A malicious user only needs to gain more control of the computing resources to cause power competition.

It is hard to define what constitutes the attributes of a "legitimate" user. In spite of the unusual behavior, identifying one as malicious (and enforce power capping) is subjective and ill-suited in an APDC.

C. *Limitation of Existing Solutions*

Current data center designs concerning power and energy management lack the capability of handling potential malicious power resource contention. Here we list three major types of recent architecture design and management strategies that have demonstrate great promise in efficiency/performance but unfortunately fail to get adapt to the insecure environment.

Distributed Management: Prior works on APDC of-ten highlights distributed energy backup (battery) [2, 3, 6, 18]. This architecture has several advantages such as improved efficiency and scalability. It also avoids a single point of failure and presents a smaller failure zone. However, the distributed energy backup does not mean attack-proof. In fact, power grab can take ad-vantage of the operation mode of distributed batteries to cause severe resource contention in Region-3.

Hierarchical Management: It is common practice for data centers to place power and energy storage systems at multiple layers and employ a hierarchical management strategy. This approach is important to ensure precise power consumption set point [18]. It also allows one to efficiently utilize the under-provisioned energy backup [17]. However, unless in Region-1, a rigid hierarchically managed data center can be easily overwhelmed.

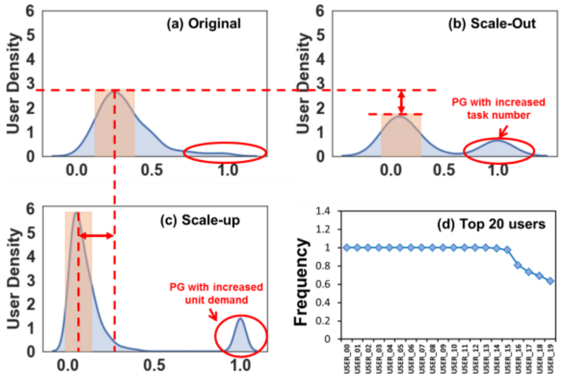


Fig.3. Feasibility of power grab from the perspective of different ways of user workload manipulation.

Fine-grained Management: Manufacturers such as Intel [19] and IBM [20] have embedded a set of Machine-Specific Registers (MSRs) into processors to monitor the power and thermal status of cores and packages. However, one cannot tell malicious users apart just by fine-grained monitoring. Power Containers [21] performs task-level power supervision and allocation to constrain load power into the cores' limits. Enforcing per-task or per-core power has no effect on preventing PG, since the attacker can choose to abuse power resource at the data center level. In addition, the key problem with PG is not capping, but fairness.

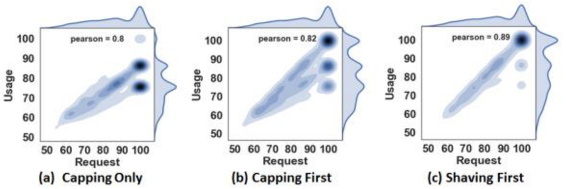
III. ANALYZING POWER GRAB

This section discusses the feasibility of power grab by investigating the impact of user-level and system level activity. We use publicly available Google cluster traces [22] to build up a stochastic model of data center. The trace contains rich information/metadata and Table II summarizes key statistical properties of the trace used. We assume a typical UPS battery that can support the full load for 15 minutes. Each server has five frequency/voltage scaling levels: 1GHz/1V, 0.9GHz/0.9V, 0.8GHz/0.8V, 0.7GHz/0.7V, 0.6GHz/0.6V. In accordance with the ability of throttling techniques and energy storages, we assume the power provisioning capacity is about 80% of total data center's nameplated power.

A. *Impact of User-Level Activities*

In today's datacenters, users are limited in interaction with the power controller. It is crucial for us to understand how user behavior affects data centers power consumption.

We examine the power behaviors of different jobs through kernel density estimation (KDE) plot within different power consumption levels. KDE estimates the probability density function. In Figures 3(a)-(c), the horizontal axis represents user's power usage normalized to the server's nameplate power. The vertical axis shows user density. The area in the figure is the total user number (normalized to 1). By looking at the KDE of the original trace shown in Figure 3-(a), we can see that there are a few power-hungry users that draw almost 100% of the nameplate power. The total power consumed by most users is less than 40%.When generating power grab, tasks that can further stress the compute resource are preferable.



(a) Capping Only (b) Capping First (c) Shaving First

Fig.4. Potential power grab user under existing data center power management schemes. Shaving: uses battery to handle power shortage; Capping: uses DVFS to limit power.

There are two major ways that one can launch malicious power attack. A scale-up power grab focuses on increasing load power demand of its current tasks as shown in Figure 3-(a). A scale-out power grab can cause peak power demand by in-creasing its task number, as shown in Figure 3-(b).

It’s clear that APDC risks increasing power capacity competition among tasks. Both scale-out power grab and scale-up power grab can change the shape of APDC. The former introduces more high-power tasks so that the relative density of low-power tasks decreases. The latter increases per-task power consumption so that other tasks have to run on a budget.

In an APDC, users that can dominate power de-mand are not unusual. In Figure 3(c), we partition the observation window into 120 equal time slots and count the top 20 power-consuming users. Our results show that there are a few users keep consuming the most amount of power. They are more likely to launch power grab.

B. *Impact of System-Level Activities*

Data center peak power controllers can exert a suppressive effect on user power demand. We further examine how existing controllers affect power grab. We compare the user’s requested power demand with the actual power allocated. As shown in Figure 4, we use the multivariate kernel density estimation (MKDE) to analyze the influence of existing power management framework to users’ request. In addition, we use Pearson correlation coefficient (PCC) as a measurement of the linear correlation between the user distribution with requested and allocated power traces. In Figures 4(a)-4(c), we calculate the Pearson correlation of the requested power with the consumed power under three representative power management schemes (detailed in Section 5) in APDC. The high Pearson value means more similar between the requested and usage. The deeper the color, the more two curves overlap. The closer the colored area is to the diagonal, the more identical the requested power is to the allocated.

Our results show that aggressive power capping has more suppressive effect. However, all the evaluated approaches have limited impact on the power behavior of power-consuming users. These users have the potential to launch a PG. To handle the disturbance that a power-hungry user may create, it is crucial to increase the resiliency of existing power management framework.

IV. CFP: CHARGES FOR POWER

We propose CFP, a power management strategy for handling power grab (PG) in an aggressively provisioned data center. CFP is also a recursive acronym for “CFP fines PG!”.

The CFP framework views non-IT and IT resource as equally important. User behavior not only affects IT resources but also non-IT resources. When non-IT resources are inadequate, the data center will adjust the user's IT resource usage accordingly, thus affecting the user experience.

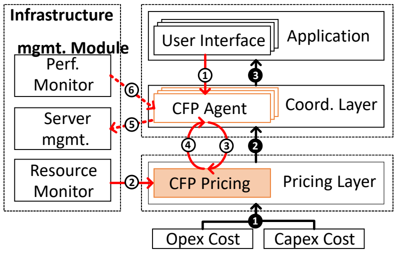


Fig.5. Overview of the CFP framework. It charges for both IT and non-IT resource usage (red arrows).

In addition, CFP emphasizes the interaction be-tween users and data centers. It is the key to avoid the blind competition for non-IT resources. CFP informs users the scarcity of power capacity and link power resource usage to their bills. Considering that the competition for non-IT resources in an APDC grows rapidly, it is reasonable to use a market-based approach to manage power budget allocation.

A. *An Overview of CFP*

CFP is a strategic power bidding scheme that can be integrated into existing designs to build a more agile and resilient power management framework. It is based on a competitive market model, where pricing is done to promote high utilization and Pareto optimal distribution. Using the competitive market terminology, data center users are consumers, intelligent power distribution units are producers, and agents are used to assist the exchange of resources in the market. Figure 5 shows the logic layers of CFP ecosystem and Figure 6 shows the overall pricing approach.

In Figure 5, the pricing layer is mainly responsible for setting the price with respect to the current users’ demand and system power supply. The power is priced to reflect supply and demand for the end of achieving fairness among the users. There is also a coordination layer between the pricing layer and user layer. It consists of numerous agents (middleware). The agents mainly perform the following tasks such as bidding involvement, purchase decisions and local power control. It also supports several power controller drivers to translate the power purchasing decision into actual power tuning activities.

It’s easy to deploy the two layers of CFP in an APDC. It’s convenient to deploy the price layer on existing cluster management system such as Google’s cluster [22]. Current kernel controlling system has monitored and maintained usage information of the bottom power infrastructure. Agents can be hosted in the loads servers or hypervisor. They are light-weight programs to profile the QoS for their supporting applications. In addition, they also run the code of deciding bidding.

Figure 6 shows the major components of CFP. There are three main questions to be answered: 1) capture the heterogeneity and variability of users; 2) dynamically match user power demand with peak power demand; and 3) motivate users to contribute to a balanced power allocation environment.

B. *SLO Runtime Profiling*

Our framework periodically profiles applications in an APDC and distills crucial power/performance variability data.

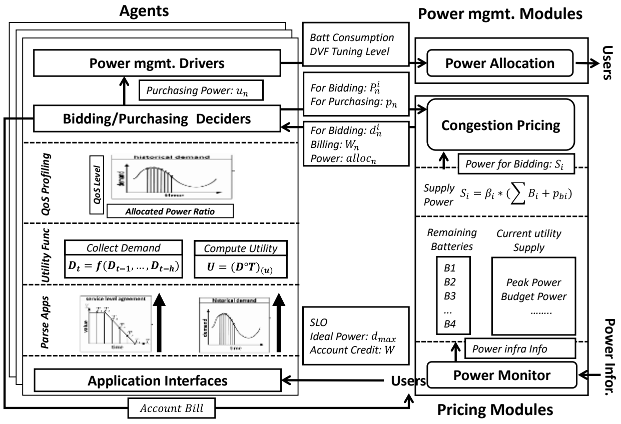
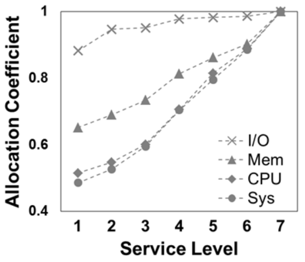


Fig.6. A market-based approach for charging PG in CFP.



Benchmark Type & Name	
Sys	*Hint, sunflow, phpbench, scikit-learn, stress-ng, xsbench, apache, pybench, cyclictest, tjbench*
CPU	*sudokut, mencoder, espeak, polybench-c, scimark2, ebizzy, graphics-magick, encode-mp3, c-ray, crafty, gcrypt, hmmer, povray*
Mem	*t-test1, stream, ramspeed*
I/O	*aio-stress, tiobench, fs-mark, iozone*

Fig.7. The performance-power curve and the benchmark evaluated in our simulation framework.

It exposes user characteristics to the data center for optimization purpose.

User requires power for supporting fundamental service quality levels. Some applications require a guarantee of certain power availability for extended duration of time, and therefore, they are likely to prefer high power budget. In contrast, some applications can easily adjust demand since their QoS is loosely related to power budget.

If we charge for prices, different users will naturally show preference to different amount of price. In economics these preferences are represented with a utility function. The utility function maps a resource amount to a real number that corresponds to a satisfaction level. In Figure 6, it provides an important link between resource occupation and user satisfaction. In this work we use power-performance profiles as the utility curves. The power-performance profile is a two-dimensional graph, as shown in Figure 7. The profile can be approximated by a piece-wise linear curve with different slopes. The slope of each linear segment rep-resents the rate at which the performance of the application degrades when the system allocates a percent-age of the demand power.

Profiles can be created for a variety of applications. New and updated profiles can be easily incorporated within the system as they become available. We characterize the profiles for workload types as shown in Figure 7. In order to draw the profile curves, we choose 30 representative tests of various application types as shown in Figure 7. We read their power consumption at different frequency level through the *turbostat* method provided by Linux kernel. We compute the average value of the tests' allocated power for a given application type. There is slight difference among the variation trend of tests belonging to different application type. For example, in Figure 7, it suggests that I/O applications are less flexible under power budget scaling.

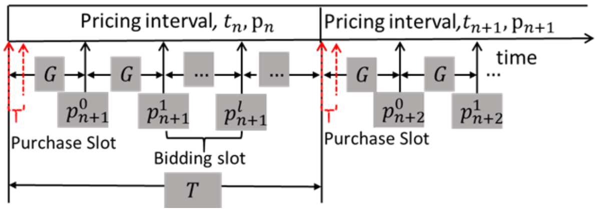


Fig.8. A tatonnement bidding process

C. *Perturb and Observe*

CFP uses a mild "perturb and observe (P&O)" approach for power allocation and customer billing. New users may come and leave; workload characteristics vary over time. In this highly dynamic environment, CFP intends to avoid frequent and abrupt power cap-ping that may disturb or disrupt user operations. We adjust the system-wide power price and observe users' responses to differentiated power capping policies. It is an iterative auction process by which power resource exchange equilibrium is expected to be achieved.

The entire data center power ecosystem can be viewed as multiple competitive markets, one market per rack. These markets operate independently and asynchronously. Consequently, this results in a decentralized system, where the failure of one rack does not necessarily cause failure of the entire economy.

In Figure 8, CFP has three power management intervals. We assume a pricing interval T, which consists of several bidding slot G. In addition, we assume the user has an equal amount of budget, W to pay for the expenses. At the beginning of each pricing interval t_n, a constant system price P_n is calculated to guide each user to make their power purchase decision. During t_n, a sequence of bidding processes is used to compute the next system price P_{n+1}. In every bidding slot, the system computes an intermediate price P_{n+1}^l. The last P_{n+1}^l in the pricing interval t_n is the system price for the whole pricing interval t_{n+1}.

For user x, the agent collects its SLO preference and the maximum required power $D_{max}^{x,l}$ at each bidding slot l. During the ding slot, $D_{max}^{x,l}$ keeps constant. According to the SLO profiling, the agent can calculate the minimum power $D_{min}^{x,l}$ for the user with regard to the minimum acceptable service level.

Bidding Process: Bidding is an interactive process that happens in the pricing interval periodically. The purpose of the bidding process is to determine the system price p_{n+1} for pricing interval t_{n+1}. The system sends back to the agent a bid price p_{n+1}^l every bidding slot, which is calculated using all the users' power demand d_n^l and the total system power supply S as shown in equation (1) and (2). $p_{n+1}^l$is an intermediate price for P_n, i.e., P_n is the last p_{n+1}^l in t_n.

$$p_{n+1}^l = p_0 * d_n^l / \alpha S \qquad (1)$$

$$d_n^l = \sum_x bid_n^{x,l} \qquad (2)$$

The power supply is the maximum available power times a constant α, where 0<α≤1. Once the agents receive the new bid price p_{n+1}^{l+1}, they compare it with the base price p_0. If it is larger than p_0, the agent reduces user's service level, otherwise, it increases the user's service level. We use τ^{l+1} to reflect the regulation process of user's SLO. In addition, there is an

TABLE III. THE REWARD AND PUNISHMENT IN CFP

	High Power	Low Power
Matching	$p_{n+1} + \sigma * p_0$	$p_{n+1} - \rho * p_0$
Mismatching	$p_{n+1} + (\sigma - \rho) * p_0$	p_{n+1}

affordable bid power $D_{afford}^{x,l+1} = W/p_{n+1}^{l}$ for user x. Then, the agent makes the next bidding decision as follows.

$$bid_n^{x,l+1} = \begin{cases} \tau^{l+1} D_{max}^{x,l+1} & \text{if } D_{afford}^{x,l+1} \geq \tau^{l} D_{max}^{x,l+1} \\ D_{afford}^{x,l+1} & \text{if } D_{min}^{x,l+1} \leq D_{afford}^{x,l+1} < \tau^{l} D_{max}^{x,l+1} \\ 0 & \text{if } D_{afford}^{x,l+1} \leq D_{min}^{x,l+1} \end{cases} \quad (3)$$

Purchase decision: There is a constant system price P_{n+1} for the whole pricing interval t_{n+1}, which is calculated in t_n. The p_{n+1}^{l} in the above bidding process is an intermediate price for P_{n+1}, i.e., P_{n+1} is the last p_{n+1}^{l} in t_n and the initial price used for t_0 is the baseline price p_0. At the beginning of the pricing interval t_{n+1} , an agent purchases power for a given user within the purchase slot. The purchased power for user x is $alloc_{n+1}^{x}$, which equals to the bid power calculated in the first bidding slot within t_{n+1}, denoted as $bid_{n+1}^{FT,x}$. The last p_{n+1}^{l} in t_n is p_{n+1}^{LT}. Thus, for an existed user x, the allocated power in t_{n+1} equals to:

$$alloc_{n+1}^{x} = bid_{n+1}^{FT,x} \quad (4)$$

Once an agent makes the purchase decision, its customer will be allocated with power budget $alloc_n^x$ during the whole pricing interval t_n no matter how load power changes. In addition, the agent uses $alloc_n^x$ to calculate the consumed power for a user within t_n.

D. Incentive Mechanism

In an APDC that charges for power, each user needs incentives to actively participate. Although CFP allows users' power demand to follow the supply via a market-based price, a malicious user can burden the other users with numerous requests (i.e., a very large d_n^l). On the other hand, it's fairer for different users to pay for differentiated price according to their demand characteristics, rather than the system price

To ensure continuous control effectiveness and fairness, CFP dynamically adjusts the price based on the reward and punishment mechanism. It fine-tunes the price of the users in accordance with their average power demand and coordination characteristics over the tatonnement pricing process. Specifically, CFP compensates normal users for contributing to power capacity savings. Attackers can increase the cost of other users at the expenditure of considerable money, ultimately cause serious imbalance between payment and performance for the other. Our compensation strategy stops it through charging extra tax while the system is under pressure or the service quality of others is near the edge of collapse.

As shown in Table III, ρ indicates the willingness of a user to regulate its power demand to facilitate the overall demand meeting the supply; σ represents how much the correlation between users' average power demand and the exceeding power over supply. CFP rewards the user with ρ>0 because it means the user coordinates its power demand along with the difference variation between demand and supply. In order to violate the provisioned power of the data center, aggressive power grab consumers must increase their request. CFP punishes these users, i.e., $\sigma > 0$ when the total demand exceeds the supply.

TABLE IV. CONFIGURATION OF OUR PROOF-OF-CONCEPT

Category	Benchmark Name
Node	4 nodes (24 cores in total) + 1 management node
CPU	Intel Xeon CPU E5-2620 v3, 6-core, 2.4G
Storage	32 GB synchronous registered memory + 1T Disk
OS	Ubuntu 14.04.3 with Linux kernel 4.4.0-31-generic

TABLE V. EVALUATED BENCHMARKS ON OUR PROTOTYPE

Type	Name	Target	Description
Normal	HPCG	CPU	Supercomputing tasks
	PHPBench	System	PHP interpreter
	Loopback	I/O	Network testing
PG	Stream	Memory	RAM testing
	Ebizzy	CPU	Web server workloads

TABLE VI. EVALUATED POWER MANAGEMENT SCHEMES

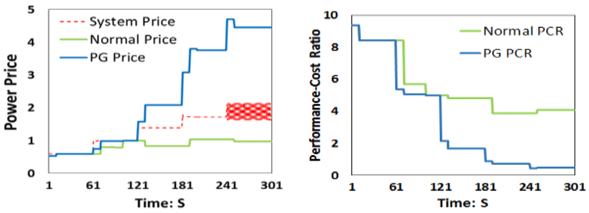
Scheme	Feature	Description
Capping	Performance scaling only	Only uses dynamic voltage and frequency scaling (DVFS) to cap power
Shaving	UPS based peak shaving	Triggers DVFS only if the UPS used for peak shaving runs out of energy
Bid	Only bidding	Only use bidding-based method to coordinate resources among servers
P&O	CFP without incentives	Only use P&O approach and congestion pricing model to manage resource
CFP	Our proposal	A resilient power capacity management framework for managing PG

V. EXPERIMENT METHODOLOGIES

We build a scaled-down testbed of CFP as well as implement a trace-based simulation environment. We analyze the impact of CFP on large scale systems using realistic Google cluster application trace.

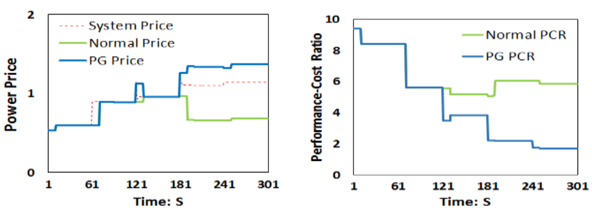
Our scaled-down CFP prototype consists of a mini server rack with four loads servers as shown in Table IV. With the ACPI, these processors support operating frequencies from 1.2GHz to 2.4 GHz at the intervals of 0.1GHz. As shown in Table V, each load server runs various tests provided by PST (Phoronix Test Suite) as the user applications. Particularly, a power-dominant attacker runs either Stream (Memory-intensive Attacker) or Ebizzy (CPU-intensive Attacker) test. The pricing center is a lightweight module implemented with C language. It runs on another server that uses Intel Celeron G1620 2-core CPU as the processing engine. Agent programs are deployed on the load servers since these programs do not affect user's application heavily. RAPL (Running Average Power Limit) interfaces allow them to constantly monitor each load server's power consumption per second. User transfers its resource requirements to the agent with a shared memory. The communication between agents and kernel pricing center is implemented by TCP/IP socket over Ethernet.

Our simulation platform consists of three parts: the input handling module, the processing elements, and the output modules. We simulate both user behavior and power supply components. We implement our CFP framework as well as representative data center power managing approaches. For simplicity, we assume homogeneous server hardware. Each server is implemented with on-core voltage and frequency scaling modules. The servers share the identical utility power supply and with one lead-acid battery per server. The battery ensures that it can maintain 15 minutes under full load. The server power, batteries and rack are dynamically monitored on a per-second basis.



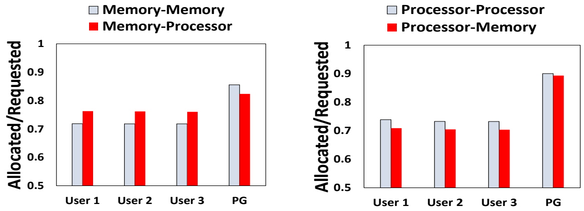
(a) Cost (b) PCR

Fig.9. System traces showing CFP with scaled-out PG.



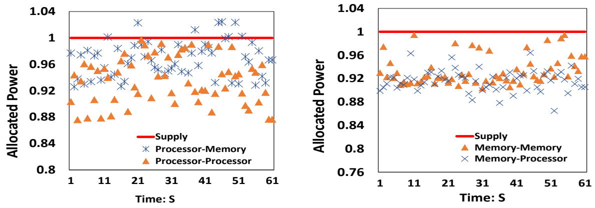
(a) Cost (b) PCR

Fig.10. System traces showing CFP with scaled-up PG.



(a) Over-estimation (b) Under-estimation

Fig.11. The impact of mis-profiling on single user.



(a) Over-estimation (b) Under-estimation

Fig. 12. The impact of mis-profiling on APDC.

We model malicious users taking advantage of several highest power-consumed users from a set of the real workloads as shown in Figure 7. We adjust the number of attackers by replacing normal users with the chosen malicious ones. We control the attack force of an attacker through increasing or decreasing its requested CPU cores.

We compare our design with other three kinds of the present power management schemes as summarized in Table VI. Among those, *Capping* is a representative peak power management technique similar to prior work [7], which only scaling down the overall servers' active power to shave peak power. *Shaving* represents a group of schemes that pay more attention on performance [1, 2]. *Bid* is an approach based on the competitive market mechanism [23]. It requires all the servers to bid for their executing power to limit the total power consumption within a safe range without considering about the tradeoff between power and quality of ser-vice. We also analyze the feasibility of the key components of the proposed CFP by comparing with *P&O*.

VI. Experiment Results

A. *System Response to PG*

We first examine how CFP framework reacts to the power grab activity by monitoring the runtime logs. Specifically, we evaluate two types of power grab activities: scale-out PG that focus on initiating additional tasks, and scale-up PG that focus on manipulating the power demand of its current tasks.

In Figures 9 and 10, we show how normal users' average power cost and performance-cost ratio (PCR) fluctuates with power grab. In Figure 9, the system receives 10 more hostile applications every minute. During the first minute, there are no power-hungry tasks. In the 60th second, a PG user increases power-consuming tasks but the data center is still able to feed all the running jobs. In the following time intervals the PG user increases power demand and the data center starts to charge both the PG user and normal users at a higher power price. In Figure 9-(a), the red line shows the base-line price that the data center used to charge normal users. Normal users (indicated by the green line) that are cooperative may choose to trades off power capacity for better price. Therefore their power price is slightly lower than the red line. As the PG user continues to increase total power demand, the punishment mechanism functions to ask for a skyrocketing price.

In Figure 9-(b), it is clear that the performance-cost ratio decreases as more tasks joining in to compete for the limited power resources. CFP intends to protect normal users, reward cooperative users, and punish greedy users. The price for normal users is close to the baseline price, while the PG user's performance-cost ratio reaches their lowest.

Increasing the power demand of existing tasks is another way to launch power grab. In this case the at-tacker must first run its workload in the lowest performance level. When the power demand of other normal users reaches certain limit, the attacker can boost its load performance and bid for more power resource. As we can see from Figure 10, such scale-up power grab normally has limited impact on other tasks. The hostile application can combine scale-out and scale-up approaches to launch a real power grab.

B. *Problems of Mis-Profiling*

The operation of CFP largely depends on its awareness of the user. If the data center mis-predicts its user's application, it can cause degraded optimization effectiveness. In Figure 11 we evaluate two scenarios, over estimation and under estimation. In the former experiment, the data center mistakenly uses the data of a CPU-intensive (sensitive to power variation) task to guide the power allocation of memory-intensive task; in the latter experiment, the data center mistakenly manages CPU-intensive workload based on a historical performance-power curve of a memory-intensive task.

We can see that in both cases, mis-profiling can lead to decreased power allocation on power grab. In Figure 10-(a), all the applications are memory-intensive. When mistakenly characterizing the PG user as CPU-intensive, the agent actually overestimates the total power de-mand. Therefore, it intends to ask for less power, resulting in more power budget allocated to normal users. In Figure 11-(b), the PG workload is CPU-intensive. When mistakenly characterizing PG as a memory-intensive user, the agent may under-estimate power demand. When the agent asks for excessive amount of power from the data center, the total baseline price can in-crease. This results in reduced actual power budget on all the users.

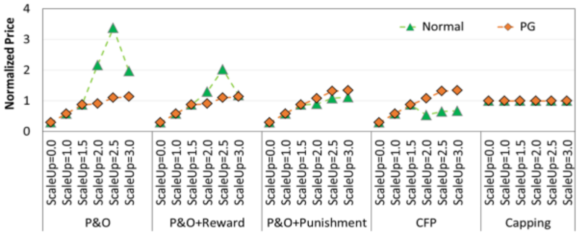
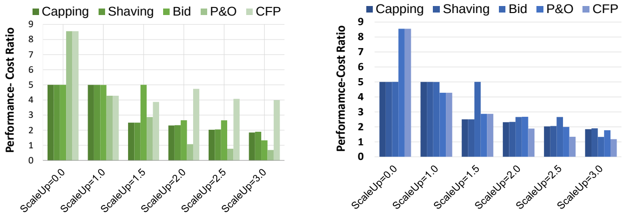


Fig. 13. The impact of reward/punishment on pricing.



(a) Normal (b) PG

Fig. 14. The impact of CFP on performance cost ratio.

In Figures 12-(a) and 11-(b) we show how mis-profiling affect the total power demand in an APDC. In both figures, the red line indicates a soft power budget. We can see that with correct profiling and control, the power demand at each timestamp is close to and al-ways below the red line. However, with mis-profiling, it is possible to result in power overload or low power utilization. Therefore, APDC that uses CFP must carefully characterize emerging and potential applications to improve the accuracy of application characterization.

C. *Impact on Price*

In Figure 13, we further evaluate the impact of the rewarding and punishment. For P&O, normal users under power grab have to greatly increase the bidding price so as to gain necessary power budget. Notably, with CFP, the price of normal user becomes much lower. It is evident that CFP allows APDC to further protect normal users from power grab.

From Figure 14 we can see that conventional designs such as Capping and Shaving maintain a constant power price in APDC. For the market-based mechanisms, the price goes up as the attack strength increases. In fact, conventional power management schemes maintain the price stability at the expense of PCR degradation. Although the average PCR decreases as malicious users create more powerful PG, CFP always maintains a better PCR for normal user.

D. *Scalability Analysis*

We compare CFP with conventional designs in terms of the average service quality of normal users and hostile users, as shown in Figure 15. When the strength of the attacker is relatively low, both CFP and the other schemes can guarantee their service quality. As the attacker increases power grab strength, conventional schemes will reduce the assigned power of all the users without any difference. With CFP, the system can still maintain desirable service quality. This is because the CFP enables the data center to support more users.

Energy efficiency is one of the key driven forces of designing APDC. Figure 16 evaluates the energy utilization of our system. It shows that CFP can make full use the power resource compared with existed schemes. As compared to the conventional power management designs, CFP makes better use of the limited power re-sources. We also measure the total throughput as the ratio of accomplished tasks number to the total tasks number. CFP is a more application-aware power management strategy since it distributes the power resource based on user feedback. Instead of running a few power hungry applications, CFP focuses on supporting many medium tasks that in total provides high throughput. Thus, CFP can support more users since the power allocation decision is more autonomous. In Figure 17, CFP can increase the throughput by 15%, depending on the scale of power grab activities.

VII. Related Work

A. *Aggressive Power Provisioning*

Power-Limited Data Centers: Over-provisioning server resources increases cost efficiency due to higher data center utilization [3]. To avoid overloading, aggressive power control strategies such as power/performance state tuning [3, 14, 24, 25], thermal/cooling optimizing [26, 27] and battery-based peak power shaving [1-4] are employed. Performance-preserving aggressive power capping framework has been deployed in the industry [15]. However, current power management frame-works mainly consider utilization and performance, overlooking malicious power resource contention issue. A sophisticated attacker can exploit the blindness of these power management schemes to mount an attack.

Fine-Grained Power Control: Also known as "power virus", some power-consuming benchmarks is designed to stress processor design [28]. There has been prior work on per-request or per-task power metering and capping [21]. Using Intel's RAPL, one can dynamically attribute the power of individual threads [29]. Although these designs allow for fine-grained power allocation, they cannot ensure a fair allocation when there are malicious loads mixed in with normal users.

B. *Market-Based Approach*

Marketed-based approaches have been used to resolve resource management issues in a variety of domains [30-33]. Researchers have developed market-based scheduling in the cloud [34-36]. However, these works mainly focus on establishing market-based model to guide the IT resource allocation, overlooking non-IT resource such as power.

A few proposals have focused on using market-based approaches to manage power resources [37-42]. For example, Wang et al. [39, 40] propose to dynamically allocate the shared power budget and the last-level cache space on chip multiprocessors. In recent papers [41, 42], researchers have proposed COOP and SpotDC which are based on supply function bidding. Nevertheless, their works mainly focus on transferring the spot capacity for improving power utilization. Differently, we focused our attention on malicious power contention resulted from oversubscribing power.

C. *Power-Related Attack*

Power & Energy Attack: Vulnerability in server power management framework has been identified recently [3, 11, 12, 13]. Prior works have discussed two challenges: energy abuse [11, 12] and power overload [3, 13]. An energy abuse-based attack mainly targets the Web application layer, with the intent of merely consuming additional server energy. Differently, we investigate a risk that arises from today's APDC. On the other aspect, a power overload-based attack aims to cause a rare,

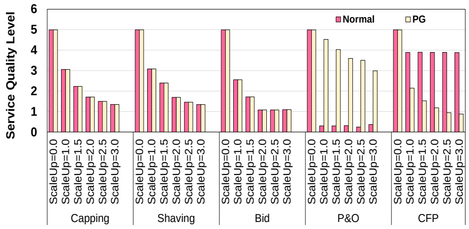


Fig.15. Variation of service levels with scaled-up PGs for different peak power management schemes.

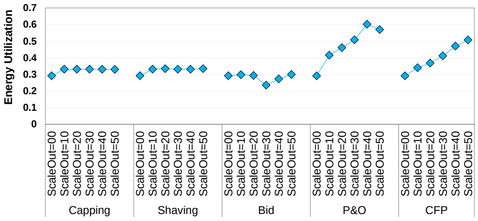


Fig. 16. Energy utilization (energy consumption divided by the aggregated maximum energy budget).

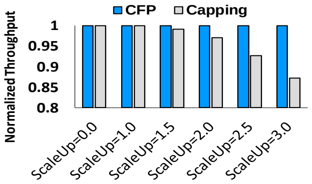


Fig. 17. CFP yields better throughput due to balanced allocation.

expensive outage. In contrast, we consider a more viable attack that focus on originating poor performance-utilization tradeoff.

Resource Availability Attack: Several papers have investigated the security issue with regard to resource contention [43, 44]. At the chip level, hardware Trojans can be used to block a network-on-chip system, causing denial of service [45]. At the system level, a resource-freeing attack (RFA) could modifying a victim VM's workloads [46]. Differently, we look at malicious load s that can deprive power resources at the facility level.

VIII. CONCLUSION

As cyber-attacks grow in sophistication and stealth, aggressively provisioned data centers are urged to be proactive in addressing threats related to power and energy. In this study we investigate power grab, a malicious act that can overwhelm the power management system of data centers. We propose CFP, an agile and resilient power management strategy that allows data center to better serve normal users in a potentially insecure environment. CFP can greatly improve the cost effectiveness of today's aggressively provisioned data center.

IX. ACKNOWLEDGEMENTS

This work is supported in part by the National Basic Research Program of China (973 Program, #2015CB352403), the National Natural Science Foundation of China (No. 61502302, 61602301, 61632017, 61702329). Corresponding author is Chao Li at Shanghai Jiao Tong University.

REFERENCES

[1] S. Govindan et al., "Benefits and Limitations of Tapping into Stored Energy for Datacenters", ISCA 2011.

[2] S. Govindan et al., "Leveraging Stored Energy for Handling Power Emergencies in Aggressively Provisioned Datacenters", ASPLOS 2012.

[3] C. Li et al., "Power Attack Defense: Securing Battery-Backed Data Centers", ISCA 2016.

[4] V. Kontorinis et al., "Managing Distributed UPS Energy for Effective Power Capping in Data Centers", ISCA 2012.

[5] S. Pelley et al., "Power routing: dynamic power provisioning in the data center", ASPLOS 2010.

[6] C. Li et al., "Enabling Datacenter Servers to Scale Out Economically and Sustainably". MICRO 2013.

[7] A. Bhattacharya et al., "The Need for Speed and Stability in Data Center Power Capping", *Sustainable Computing: Informatics and Systems*, Vol. 3, Issue 3, 2013.

[8] Ponemon Institute. "Cost of Denial of Service Attacks", *Data Center Performance Benchmark Series*, 2016.

[9] Cisco. "Defeating DDOS Attacks", 2014.

[10] D.H. Woo and H.H. Lee. "Analyzing Performance Vulnerability due to Resource Denial-of-Service Attack on Chip Multiprocessors", *CMP-MSI Workshop*, 2007.

[11] F. Palmieri et al., "Energy-Oriented Denial of Service Attacks: An Emerging Menace for Large Cloud Infrastructures", *The Journal of Supercomputing*, Volume 71, Issue 5, pp 1620–1641.

[12] Z. Wu et al., "On Energy Security of Serer System", *IEEE Transactions on Dependable and Secure Computing*, Volume 9, No. 6, 2012.

[13] Z. Xu et al., "Power Attack: An Increasing Threat to Data Centers", NDSS 2014.

[14] X. Fan et al., "Power Provisioning for a Warehouse-Sized Computer", ISCA 2007.

[15] Q. Wu et al., "Dynamo: Facebook's Data Center-Wide Power Management System", ISCA 2016.

[16] D. Wang et al., "ACE: Abstracting, Characterizing and Exploiting Datacenter Power Demands", IISWC 2013.

[17] D. Wang et al., "Under-provisioning Backup Power Infrastructure for Datacenters", ASPLOS 2014.

[18] L. Liu et al., "HEB: Deploying and Managing Hybrid Energy Buffers for Improving Datacenter Efficiency and Economy", ISCA 2015

[19] "Intel® 64 and IA-32 Architectures Software Developer's Manual", Volume 3B: System programming Guide, Part 2, 2011, Intel

[20] An Introduction to POWER8 Processor. http://www.idh.ch/IBM_TU_2013/Power8.pdf.

[21] K. Shen et al., "Power Containers: An OS Facility for Fine-Grained Power and Energy Management on Multicore Servers", ASPLOS 2013.

[22] Google Cluster-Usage Traces: Format + Schema. https://drive.google.com/file/d/0B5g07T_gRDg9Z0lsSTEtTWtpOW8/view.

[23] M. Islam et al., "A Market Approach for Handling Power Emergencies in Multi-tenant Data Center", HPCA 2016.

[24] Y. Chen et al., "Managing Server Energy and Operational Costs in hosting Centers", SIGMETRICS 2005.

[25] D. Meisner et al., "Power Management of Online Data-Intensive Services", ISCA 2011.

[26] W. Zheng, et al., "TECfan: Coordinating Thermoelectric Cooler, Fan, and DVFS for CMP Energy Optimization", IPDPS, 2016.

[27] X. Gao et al., "Reduced Cooling Redundancy: A New Security Vulnerability in a Hot Data Center", NDSS 2018.

[28] K. Ganesan et al., "System-level Max Power (SYMPO)-A systematic approach for escalating system-level power consumption using synthetic benchmarks", PACT 2010.

[29] Y. Zhai et al., "HaPPy: Hyperthread-aware Power Profiling Dynamically". USENIX ATC, 2014.

[30] M. Guevara et al., "Navigating Heterogeneous Processors with Market Mechanisms", HPCA 2013.

[31] S. Ha et al., "Tube: Time-dependent Pricing for Mobile Data," SIGCOMM 2012.

[32] B. HomChaudhuri, and K. Manish. "Market based Allocation of Power in Smart Grid", *American Control Conference* (ACC), 2011. IEEE, 2011.

[33] E. Fulp, et al., "Paying for QoS: An Optimal Distributed Algorithm for Pricing Network Resources", IWQoS 1998.

[34] B. Pittl et al., "A Negotiation-based Resource Allocation Model in IaaS-Markets", *Utility and Cloud Computing* (UCC), 2015.

[35] A. Byde et al., "Market-based Resource Allocation for Utility Data Center", HP Lab, Technical Report HPL-2003-188, 2003.

[36] M. Macias, and G, Jordi. "A Risk-based Model for Service Level Agreement Differentiation in Cloud Market Providers", DAIS 2014.

[37] B. Lubin et al., "Expressive Power-Based Resource Allocation for Data Centers", IJCAI 2009.

[38] M. Perninge, and E. Robert. "Frequency Control in Power Systems Based on a Regulating Market", *IEEE Transactions on Control Systems Technology*, 2018.

[39] X. Wang, and M. José F. "ReBudget: Trading off Efficiency vs. Fairness in Market-based Multicore Resource Allocation via Runtime Budget Reassignment", ASPLOS 2016

[40] X. Wang, and M. José F. "XChange: A Market-based Approach to Scalable Dynamic Multi-Resource Allocation in Multicore Architectures", HPCA 2015.

[41] M. Islam et al., "A Market Approach for Handling Power Emergencies in Multi-tenant Data Center", HPCA 2016.

[42] M. Islam et al., "A Spot Capacity Market to Increase Power Infrastructure Utilization in Multi-Tenant Data Centers", SIGMETRICS 2017.

[43] D. Grunwald and S. Ghiasi. "Microarchitectural Denial of Service: Insuring Microarchitectural Fairness", MICRO 2002.

[44] D.H. Woo and H.H. Lee. "Analyzing Performance Vulnerability due to Resource Denial-of-Service Attack on Chip Multiprocessors", *CMP-MSI* 2007.

[45] T. Boraten and A. Kodi. "Mitigation of Denial of Service Attack with Hardware Trojans in NoC Architectures", IPDPS 2016.

[46] V. Varadarajan et al., "Resource-Freeing Attacks: Improving Your Cloud Performance (at Your Neighbor's Expense)", CCS 2012