Enabling Reliable Data Center Demand Response via Aggregation

Luyao Niu
Oklahoma State University,
Stillwater, OK
luyao.niu@okstate.edu

Yuanxiong Guo
Oklahoma State University,
Stillwater, OK
richard.guo@okstate.edu

ABSTRACT

Although data centers are recognized as promising resources for demand response (DR), it is not easy for them to participate into DR programs due to their unreliable DR capacities. The unreliability mainly comes from their random workload arrivals. In this paper, we study how to enable reliable data center DR. We focus on the scenario that independent data centers participate into capacity bidding program (CBP) in which they need to sign forward contracts to commit the amount of power reduction during the DR event. We show that due to the uncertainty of DR capacity in real time, it is risky for the data centers to sign contracts in advance. Such risky behaviors are adverse to the profitability of data centers when providing DR resources. Inspired by the intuition that aggregation can reduce uncertainty, we propose that data centers cooperate with others and sign the forward contract collectively based on their aggregated DR capacity to mitigate the uncertainty of DR capacity. A coalitional game is used to model the cooperation among the data centers. We further design a payoff allocation to split the profit generated via cooperation fairly, guaranteeing that no coalition has the incentive to deviate. In addition, we show that the proposed payoff allocation captures the marginal contribution of each data center and is efficient. Finally, trace driven simulation results are presented to demonstrate the effectiveness of the proposed approach. The results show that participating into CBP collaboratively leads to a win-win situation where the data centers obtain higher profits and the utility company gets more reliable DR resources from data centers.

CCS Concepts

•Hardware → Enterprise level and data centers power issues; •Theory of computation → Solution concepts in game theory;

Keywords

Data center, demand response, reliability, coalitional game

1. INTRODUCTION

Our power grid nowadays is experiencing higher penetration of renewable energy. The intermittent and fluctuating renewables introduce a crucial challenge to the balance between supply and demand of the power system. Demand response (DR), which is an important component of the future smart grid [1, 3, 6], is an elegant and economic feasible solution to address the challenge by controlling the demand side to maintain the power balance. Additionally, several benefits can be achieved by DR. First, the consumers can receive incentive payments or reduce electricity bills if they respond to the programs. Second, the reliability of the power system can be improved, which reduces the risk of outages [7]. Furthermore, the benefit is also market-wide. The system capacity is improved, and the electricity price is expected to decrease through DR. Finally, DR is also beneficial to the electricity market by improving the market performance and reducing the volatility of electricity price. Considering all these attractive and important features, more and more attentions are paid to DR. According to the report from ISO-NE [2], the growth of DR resources in New England from 2003 to 2010 is more than 2000 MW.

Data centers, which serve as the backbone of the modern economy, are valuable and promising DR resources [8, 13, 40]. First and foremost, the flexibility of data centers, which has been explored by previous works [18, 19, 38, 42], provides the potential to participate into DR programs. The flexibility can be obtained using either workload management techniques, or on-site generation and energy storage. Second, data centers are usually large loads in the power grid, and their energy consumption is growing fast [23, 27]. For instance, the electricity consumption of data centers in US was about 91 billion kilowatt-hour in 2013, leading to an estimated $9 billion electricity bill, and the annual growth rate of the electricity consumption of data centers is expected to be more than 10% [27]. In a word, data centers are large, yet flexible loads in the power system. With their participation in the DR programs, in the ideal case, we can reach a desirable win-win situation where the data centers can obtain more financial benefits and the sustainability of the power system can be improved.

Even though data centers are promising DR resources, it is far from easy for them to participate into the existing DR programs. One important reason is that the DR capacity of a data center highly depends on its workload arrival,

which is uncertain. In some DR programs, the amount of DR capacity needs to be committed in advance. To participate into such DR programs, data centers as DR resources are required to either bid into a forward market or sign a forward contract [4, 5]. The purpose is for ease of making decisions on unit commitment or generation reserve for utility companies [33], and reducing the potential risks effectively [11, 35]. However, these actions in advance are potentially risky to the data centers. From the perspective of data centers, they may incur heavy penalties if they fail to deliver the committed amount of DR due to the unexpected rise of their workload arrivals. Such uncertain responses in real time will potentially drive data centers away from the DR programs because data centers care more about the DR programs or events in which they can achieve benefits [6]. From the perspective of the system operator, the unreliability of data center DR capacities can make the unit commitment or generation reserve decisions inefficient since the actual load reduction depends on the realized DR capability [6]. Therefore, we are faced a challenge of how to make data centers reliable DR resources, especially when they participate into the DR programs which require planning in advance. If this challenge cannot be addressed properly, the desired win-win situation mentioned above will unfortunately degrade to a fail-fail situation.

Considering the potential of data centers to provide DR resources, several priori works have been done. They could be roughly divided into two categories. The first category focuses on how to extract the flexibilities of data centers, while the second category studies DR market or mechanism design for data centers. Both of them are important to the development of data center DR. Typical works in the first category include workload shifting, geographical workload balancing and so on [19, 20]. The works in the second category can be further divided into two streams. The first stream studies pricing based mechanisms such as coincident peak pricing [21, 22], while the second stream mainly considers incentive based DR mechanisms such as auction or supply function bidding based mechanism design [9, 17, 46]. However, most of these works do not consider the uncertainties of data center DR capacities when participating into a DR program that needs committed DR capacity in advance. Therefore, the reliability of the data center DR when participating into such DR programs has not been addressed in the existing literature.

In this paper, we focus on enabling reliable data center DR by mitigating the uncertainty of data center DR capacity when participating into the capacity bidding program (CBP), which is a typical example that requires planning in advance [4, 5]. To participate into the CBP, a data center, whose objective is to maximize its profit, needs to sign a forward contract which defines the committed amount of DR. Due to the uncertain response in real time caused by the random workload arrival, the data center might be penalized when it fails to deliver the committed amount of DR. An intuition to reduce the uncertainty is that multiple independent data centers can cooperate with others to exploit the statistical diversities of their workload arrivals and sign the forward contract collectively based on their aggregated DR capacity. To study the cooperation process, we use the coalitional game theory to model the cooperation among data centers. We show that cooperation can increase the total profit obtained through participating into the CBP, and the data centers will form into the grand coalition to maximize the total expected profit. Furthermore, we design a payoff allocation to divide the profit generated by cooperation such that it stabilizes the game, satisfies fairness and is computationally efficient.

To summarize, our contributions are as follows:

- We focus on how to enable reliable data center DR and manage their participation into the CBP. To the best of our knowledge, this is the first work on the management of data centers in CBP when considering the uncertainty of data center DR capacity.

- To mitigate the unfavorable uncertainty of data center DR capacity, we propose that multiple independent data centers participate into the CBP collaboratively and sign the forward contract collectively based on the aggregated DR capacity. A coalitional game based approach is proposed to analyze the cooperation process among data centers.

- The benefits obtained through collaboration are analyzed and quantified. We further design a payoff allocation to distribute the total profit generated through collaboration so that the game is stabilized. Additionally, we show the proposed payoff allocation captures the marginal contribution of each data center, satisfies efficiency and has low computation complexity.

- We have conducted simulations using the real world trace data to show the effectiveness of the proposed approach.

2. RELATED WORK

As we have mentioned before, the works focusing on data center DR can be roughly divided into two streams. The first stream studies how to extract the flexibilities of data centers, while the second stream focuses on DR market or mechanism design.

Workload management approaches are widely studied to extract the flexibilities of data centers. There are many works focus on power proportional system design [18, 39]. Workload scheduling, which exploits the temporal diversity in renewable energy and workload intensity, is another technique to extract data center flexibilities [19, 43]. In addition, energy storage is also used to extract data center flexibilities [14, 15, 37]. Furthermore, geographical load balancing is investigated to exploit the spatial diversity among multiple geographically distributed data centers [20, 30, 36]. Geographical load balancing can not only be used to extract the flexibilities of each data center [20, 30, 36], but also be helpful to maintain the reliability of the power system such as preventing overflow on the transmission lines [12, 24, 25].

Another stream of works focuses on DR market or mechanism design. On one hand, some works propose to use pricing based mechanisms, i.e., data centers are exposed to price signals issued by the system operator and will respond to them [17, 21, 22, 36]. On the other hand, some efforts are devoted to incentive based mechanisms, i.e., data centers bid their marginal costs or DR capabilities to the system operator [9, 45, 46]. Emergency DR (EDR) programs, which mainly target at load reduction under emergent scenarios such as extreme weather events, are typical incentive based DR programs and have been studied by some works [10, 44].

CBP also belongs to incentive based DR programs. The difference is that existing works do not consider the DR program that requires forward contract bidding and therefore cannot address the uncertainty of data center DR capacity. Our work focuses on how to make data centers reliable resources when participating into CBP, and thus is complementary to the existing literature on data center DR.

3. SYSTEM MODEL

In this section, we introduce the system model used in this paper including the data center DR capacity model and the CBP model.

3.1 Data Center DR Capacity Model

Consider a set of independent data centers $\mathcal{N} = \{1, 2, \cdots, N\}$ belonging to different entities participating into the CBP issued by a utility company. For each data center $i \in \mathcal{N}$, it owns M_i homogeneous servers whose idle power and peak power are P_i^{idle} and P_i^{peak}, respectively. According to the power consumption model in [16], when data center i keeps x servers active, its average power consumption can be represented as

$$P_i(x) = x \left[P_i^{\text{idle}} + U_i(P_i^{\text{peak}} - P_i^{\text{idle}}) \right], \qquad (1)$$

where U_i is the average CPU utilization level. In the following, we assume an M/GI/1 Processor Sharing (PS) queue is adopted at each server [20]. If the workload arrival rate is λ_i and the service rate per server is μ_i, then U_i can be calculated as $U_i(x) = \lambda_i/(\mu_i x)$. We assume that when data center i does not participate into the CBP, it keeps all servers active and uniformly distributes the workload among its servers. In this case, its average power consumption can be represented as

$$P_i^{\text{ref}} = M_i \left[P_i^{\text{idle}} + U_i^{\text{ref}}(P_i^{\text{peak}} - P_i^{\text{idle}}) \right], \qquad (2)$$

where $U_i^{\text{ref}} = \lambda_i/(\mu_i M_i)$ is the average CPU utilization level. Even though there are multiple techniques for power reduction such as workload redistribution and CPU frequency adjustment, we assume each data center i reduces its power consumption only by turning off the unused servers. If m_i servers are turned off by data center i during the DR period, then its power consumption P_i' can be calculated as

$$P_i' = (M_i - m_i) \left[P_i^{\text{idle}} + U_i'(P_i^{\text{peak}} - P_i^{\text{idle}}) \right], \qquad (3)$$

where $U_i' = \lambda_i/(\mu_i(M_i - m_i))$ is the average CPU utilization level. Hence, the power reduction ΔP_i from data center i by turning off m_i servers can be calculated as

$$\Delta P_i = P_i^{\text{ref}} - P_i' = m_i P_i^{\text{idle}}. \qquad (4)$$

For each data center i, the maximum number of servers that can be turned off m_i is limited by the quality of service (QoS) requirement. To characterize QoS, we focus on the processing delay of each data center i, which is represented as

$$T_i = \frac{1}{\mu_i - \frac{\lambda_i}{M_i - m_i}} \leq T_i^{\text{max}}, \qquad (5)$$

where T_i^{max} is the maximum average delay of the workload that can be tolerated at data center i. From the QoS constraint, we have the following inequalities

$$0 \leq m_i \leq M_i - \frac{\lambda_i}{\mu_i - \frac{1}{T_i^{\text{max}}}}. \qquad (6)$$

To capture the associated non-IT power reduction such as cooling when turning off servers, we adopt power usage effectiveness (PUE) γ_i defined as the ratio of the total power consumption to the IT power consumption in data center i. Based on (4) and (6), the DR capacity D_i of data center i can be represented as a function of the workload arrival rate λ_i as follows:

$$D_i(\lambda_i) = \gamma_i P_i^{\text{idle}} \left(M_i - \frac{\lambda_i}{\mu_i - \frac{1}{T_i^{\text{max}}}} \right). \qquad (7)$$

Note that the workload λ_i in the above equation is random. Therefore, the DR capacity $D_i(\lambda_i)$ of each data center i is also random.

3.2 Capacity Bidding Program Model

In CBP, the data centers are required to sign contracts to commit the load reduction in advance. For instance, the CBP offered by PG&E [4] provides the day-of and day-ahead operations. The participants are incentivized based on their commitments and are penalized when failing to deliver the committed amount of load reduction. In the following, we model CBP as a two-settlement market consisting of a forward market and a spot market.

First we consider the forward market. In the forward market, before the DR capacity is realized, a forward contract C_i is offered by each data center i based on its ex-ante belief in its DR capacity $D_i(\lambda_i)$. The forward contract C_i represents the committed amount of DR from data center i. Let p denote the price of the forward contract $C_i, \forall i$. The revenue obtained by each data center i in the forward market is pC_i. Here, the price p is defined in CBP, and is known to the data centers beforehand.

Next we consider the spot market. For each data center i, we propose the following procedure to settle the imbalance between the real time DR capacity $D_i(\lambda_i)$ and its forward contract C_i. The negative imbalance is settled at a penalty rate a, i.e., data center i with shortfall on DR capacity in real time suffers a penalty $a[C_i - D_i(\lambda_i)]^+$. For each unit amount of positive imbalance, a reward rate b is issued, i.e., data center i with surplus gains an extra $b[D_i(\lambda_i) - C_i]^+$ reward from the system operator. In practice, b may be as low as 0. This imbalance settlement setting is reasonable from the perspective of the market operator. For example, if a data center is not able to reach the promised DR capacity, then the system operator has to recruit other resources from the spot market to guarantee the performance of CBP. The cost incurred by the system operator during the recruitment should be paid by the data center with shortfall. Here, the penalty rate and the reward rate are constants defined in the CBP program, and are known to the data centers. Furthermore, we assume that $a \geq p \geq b \geq 0$. Here, the nonnegative assumption is to guarantee that any imbalance is not profitable, i.e., deviating from the forward contract impairs the profitability of the data center. The assumption that the reward rate is not higher than the penalty rate is to prevent the arbitrage during the DR event.

4. PARTICIPATING INTO CBP BY AGGREGATING DR CAPACITIES

According to the CBP model described above, any data center with negative imbalance in real time will be penalized. Since data centers are profit maximizing, the poten-

tial heavy penalty results in conservative or even no DR commitment in advance. To properly manage data centers participating into CBP, we propose to mitigate the uncertainty in real time response by encouraging data centers to participate into CBP collaboratively based on their aggregated DR capacity. The inspiration is the widely accepted intuition that aggregation can reduce uncertainty. In the following, we study how the data centers participate into CBP by aggregating their DR capacities.

4.1 Coalitional Game Model

In the following, we study how the data centers form into coalition and sign the forward contract collectively based on their aggregated DR capacity. We assume that all data centers are price takers and profit maximizing.

Consider S data centers form into a coalition $\mathcal{S}$. Since the aggregated DR capacity of coalition $\mathcal{S}$ is determined by the amount of DR capacity contributed by each data center in coalition $\mathcal{S}$, we define an aggregation level vector $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_S]$, where each element $0 \leq \alpha_i \leq 1$, $\forall i$ represents the aggregation level of data center i. Then it follows that the amount of DR contributed by data center i is $\alpha_i D_i(\lambda_i)$. The aggregated DR capacity of coalition $\mathcal{S}$ with aggregation level vector $\boldsymbol{\alpha}$ can be represented as $\sum_{i=1}^{S} \alpha_i D_i(\lambda_i)$, $\forall \mathcal{S} \subseteq \mathcal{N}$. Without loss of generality, in the following we let the dimension of the aggregation level vector $\boldsymbol{\alpha}$ always equal to its associated coalition size $|S|$. To capture how the aggregated DR capacity is determined by the workload arrivals, according to (7), we first define a modified aggregated workload power consumption $A_{\mathcal{S}}$ and a constant $B_{\mathcal{S}}$ associated with each coalition $\mathcal{S}$ as follows:

$$A_{\boldsymbol{\alpha},\mathcal{S}}(\boldsymbol{\lambda}) = \sum_{i=1}^{S} \alpha_i \gamma_i P_i^{\text{idle}} \frac{\lambda_i}{\mu_i - \frac{1}{T_i^{\max}}}, \qquad (8)$$

$$B_{\boldsymbol{\alpha},\mathcal{S}} = \sum_{i=1}^{S} \alpha_i \gamma_i P_i^{\text{idle}} M_i, \qquad (9)$$

where $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \cdots, \lambda_S]$ is the workload vector for coalition $\mathcal{S}$. Then the aggregated DR capacity of coalition $\mathcal{S}$ with aggregation level $\boldsymbol{\alpha}$ can be calculated as

$$\sum_{i=1}^{S} \alpha_i D_i(\lambda_i) = B_{\boldsymbol{\alpha},\mathcal{S}} - A_{\boldsymbol{\alpha},\mathcal{S}}(\boldsymbol{\lambda}). \qquad (10)$$

In the following, we use the coalitional game theory to model the cooperation process among the data centers. The coalitional game theory is a tool to analyze how the data centers are incentivized under the cooperative setting [26]. We model the cooperation process as a coalitional game using the characteristic form $(\mathcal{N}, v_{\boldsymbol{\alpha}})$ defined as follows:

- The set of data centers $\mathcal{N}$ is the set of players in the coalitional game.

- The characteristic function $v_{\boldsymbol{\alpha}}$ associated with any subset $\mathcal{S}$ of $\mathcal{N}$ with aggregation level $\boldsymbol{\alpha}$ is defined as:

$$v_{\boldsymbol{\alpha}}(\mathcal{S}) = \max_{C_{\boldsymbol{\alpha},\mathcal{S}}} pC_{\boldsymbol{\alpha},\mathcal{S}} + \mathbb{E}\left[b[(B_{\boldsymbol{\alpha},\mathcal{S}} - A_{\boldsymbol{\alpha},\mathcal{S}}) - C_{\boldsymbol{\alpha},\mathcal{S}}]^+\right]$$

$$- \mathbb{E}\left[a[C_{\boldsymbol{\alpha},\mathcal{S}} - (B_{\boldsymbol{\alpha},\mathcal{S}} - A_{\boldsymbol{\alpha},\mathcal{S}})]^+\right], \forall \mathcal{S} \subseteq \mathcal{N} \quad (11)$$

where $C_{\boldsymbol{\alpha},\mathcal{S}}$ is the forward contract of coalition $\mathcal{S}$ with aggregation level $\boldsymbol{\alpha}$. The characteristic function (11) describes the expected profit maximization problem of any coalition $\mathcal{S}$ under any aggregation level.

To solve the expected profit maximization problem above, the coalition needs to properly select a forward contract. For any coalition $\mathcal{S}$, we show that the optimal solution $C_{\boldsymbol{\alpha},\mathcal{S}}^*$ can be obtained in a closed form. Before that, we first introduce some required notations. Denote the cumulative distribution function (CDF) of $A_{\boldsymbol{\alpha},\mathcal{S}}$ as $F_{\boldsymbol{\alpha},\mathcal{S}}(z)$, which is defined as

$$F_{\boldsymbol{\alpha},\mathcal{S}}(z) = \Pr\left(A_{\boldsymbol{\alpha},\mathcal{S}(\lambda)} \leq z\right). \qquad (12)$$

Given the historical workload information, the CDF of the aggregated workload power consumption can be easily constructed. It follows that the quantile function $F_{\boldsymbol{\alpha},\mathcal{S}}^{-1}(\tau)$ is defined as

$$F_{\boldsymbol{\alpha},\mathcal{S}}^{-1}(\tau) = \inf\left\{z \in \left[A_{\boldsymbol{\alpha},\mathcal{S}}^{\min}, A_{\boldsymbol{\alpha},\mathcal{S}}^{\max}\right] : \tau \leq F_{\boldsymbol{\alpha},\mathcal{S}}(z)\right\}, \quad (13)$$

where $A_{\boldsymbol{\alpha},\mathcal{S}}^{\min}$ and $A_{\boldsymbol{\alpha},\mathcal{S}}^{\max}$ are the lower and upper bound of the modified aggregated workload power consumption $A_{\boldsymbol{\alpha},\mathcal{S}}$ (9), respectively. Given the CDF of the modified aggregated workload power consumption $A_{\boldsymbol{\alpha},\mathcal{S}}$ of coalition $\mathcal{S}$, the optimal forward contract $C_{\boldsymbol{\alpha},\mathcal{S}}^*$ can be obtained in a closed form as shown in the following theorem.

THEOREM 1. *The optimal forward contract for any coalition $\mathcal{S}$ with aggregation level $\boldsymbol{\alpha}$ can be obtained as follows:*

$$C_{\boldsymbol{\alpha},\mathcal{S}}^* = B_{\boldsymbol{\alpha},\mathcal{S}} - F_{\boldsymbol{\alpha},\mathcal{S}}^{-1}(\tau^*), \quad where \quad \tau^* = \frac{a - p}{a - b}. \quad (14)$$

PROOF. Note that (11) can be rewritten as

$$v_{\boldsymbol{\alpha}}(\mathcal{S}) = pC_{\boldsymbol{\alpha},\mathcal{S}} - a \int_{B_{\boldsymbol{\alpha},\mathcal{S}} - C_{\boldsymbol{\alpha},\mathcal{S}}}^{A_{\mathcal{S}}^{\max}} (C_{\boldsymbol{\alpha},\mathcal{S}} - z) f_{\boldsymbol{\alpha},\mathcal{S}}(z)\,\mathrm{d}z$$

$$+ b \int_{A_{\boldsymbol{\alpha},\mathcal{S}}^{\min}}^{B_{\boldsymbol{\alpha},\mathcal{S}} - C_{\boldsymbol{\alpha},\mathcal{S}}} (z - C_{\boldsymbol{\alpha},\mathcal{S}}) f_{\boldsymbol{\alpha},\mathcal{S}}(z)\,\mathrm{d}z, \quad (15)$$

where $f_{\boldsymbol{\alpha},\mathcal{S}}$ is the probability density function of the aggregated workload power consumption of coalition $\mathcal{S}$ with aggregation level $\boldsymbol{\alpha}$. Using Leibniz integral rule and the first-order optimality condition, we have

$$p - a\left(1 - F_{\boldsymbol{\alpha},\mathcal{S}}(B_{\boldsymbol{\alpha},\mathcal{S}} - C_i^*)\right) + bF_{\boldsymbol{\alpha},\mathcal{S}}(B_{\boldsymbol{\alpha},\mathcal{S}} - C_i^*) = 0. \quad (16)$$

Then it is straightforward to see that Theorem 1 holds. $\square$

By replacing the forward contract $C_{\boldsymbol{\alpha},\mathcal{S}}$ in (15) with the optimal solution (14), the optimal expected profit for coalition $\mathcal{S}$ with aggregation level $\boldsymbol{\alpha}$ is represented as

$$v_{\boldsymbol{\alpha}}(\mathcal{S}) = a \int_{\tau^*}^{1} F_{\boldsymbol{\alpha},\mathcal{S}}^{-1}(z)\,\mathrm{d}z + b \int_{0}^{\tau^*} F_{\boldsymbol{\alpha},\mathcal{S}}^{-1}(z)\,\mathrm{d}z. \qquad (17)$$

4.2 Grand Coalition Formation

Intuitively, the uncertainty of the aggregated DR capacity decreases as more data centers join into the coalition. In the following, we prove that signing the forward contract as a single DR provider can increase the total expected profit of the coalition. Then, it is straightforward to see that it is to the joint benefit of the data centers to form the grand coalition $\mathcal{N}$ and participate with their full DR capacities. First, we characterize the optimal expected profit defined in (11).

LEMMA 1. *The optimal expected profit $v_{\boldsymbol{\alpha}}(\mathcal{S})$ of any coalition $\mathcal{S}$ with aggregation level vector $\boldsymbol{\alpha}$ has the following properties:*

1. *Positive Homogeneity: For any nonnegative scalar ε, we have that $v_{\varepsilon\boldsymbol{\alpha}}(\mathcal{S}) = \varepsilon v_{\boldsymbol{\alpha}}(\mathcal{S})$.*

2. *Supper Additivity: For any two disjoint coalitions $\mathcal{S}_1$ and $\mathcal{S}_2$ whose aggregation vectors are $\boldsymbol{\alpha}$, we have that $v_{\boldsymbol{\alpha}}(\mathcal{S}_1 \cup \mathcal{S}_2) \geq v_{\boldsymbol{\alpha}}(\mathcal{S}_1) + v_{\boldsymbol{\alpha}}(\mathcal{S}_2)$.*

3. *Nondecreasing over $\boldsymbol{\alpha}$: For any two aggregation vectors, if $\boldsymbol{\alpha} \succeq \boldsymbol{\alpha}'$, then $v_{\boldsymbol{\alpha}}(\mathcal{S}) \geq v_{\boldsymbol{\alpha}'}(\mathcal{S})$.[1]*

PROOF. See Appendix A for detailed proof. $\square$

According to Lemma 1, we have the following observations. First, using the supper additivity property, it follows that if there are two disjoint coalitions $\mathcal{S}, \mathcal{S}' \subset \mathcal{N}$ whose aggregation level are $\boldsymbol{\alpha} = \mathbf{1}$, then we have

$$v_{\mathbf{1}}(\mathcal{S}) + v_{\mathbf{1}}(\mathcal{S}') \leq v_{\mathbf{1}}(\mathcal{S} \cup \mathcal{S}'). \quad (18)$$

The inequality above indicates that forming into one coalition and signing the forward contract collectively based on the aggregated DR capacity can generate larger profit. Therefore, any two disjoint coalitions will merge into a single coalition with a larger size and participate into CBP as a single DR provider to maximize the expected profit. Since the grand coalition is the one with the largest size, no matter what initial partition is formed in $\mathcal{N}$, the data centers will form the grand coalition $\mathcal{N}$ to maximize the expected profit. Second, since the expected profit is nondecreasing over the aggregation level vector $\boldsymbol{\alpha}$, then for any coalition, its optimal expected profit can be obtained by setting $\boldsymbol{\alpha} = \mathbf{1}$, i.e., all data centers in coalition $\mathcal{S}$ fully participate into the DR program. Combining the above two observations, we conclude that the data centers can generate the maximum expected profit by forming into the grand coalition $\mathcal{N}$ and participating into the CBP with fully aggregated DR capacity $\sum_{i=1}^{N} D_i(\lambda_i)$. For simplicity, in the following we use $v(\mathcal{N})$ to represent the expected profit obtained by the grand coalition with aggregation level $\boldsymbol{\alpha} = \mathbf{1}$.

5. PAYOFF ALLOCATION

In Section 4, we have shown that the data centers will form into the grand coalition $\mathcal{N}$ in our game. We have adopted $v(\mathcal{N})$ to describe the optimal total amount of expected profit received by the grand coalition, however, we have not explicitly specified the payoff obtained by each data center in the coalition. In the following, we investigate how to fairly distribute the total profit among the data centers in the grand coalition. Note that the data centers are assumed to be rational in our game. The intuitive solution that uniformly distributes is not applicable since some data centers might be able to obtain higher profit if it participates into CBP individually. Therefore, the payoff allocation should fairly distribute the profit among the data centers so as to stabilize the grand coalition, i.e., no deviation is profitable. In the coalitional game theory, the set of such allocations is called the core of the game, which is defined as [28]

$$\mathcal{C} = \left\{ \sum_{i\in\mathcal{N}} \pi_i = v(\mathcal{N}), \sum_{i\in\mathcal{N}} \pi_i \geq v(\mathcal{S}), \; \forall \mathcal{S} \subseteq \mathcal{N} \right\}, \quad (19)$$

[1] The operator $\succeq$ denotes element wise vector comparison, i.e., if we have $\alpha_i \geq \alpha'_i$, $\forall i \in [1, S]$, then $\boldsymbol{\alpha} \succeq \boldsymbol{\alpha}'$.

where π_i is the payoff obtained by data center i in the grand coalition. The Shapley value [32, 34] and the nucleolus [29] are two solution concepts that are widely adopted in the existing literature, however, they are not applicable in our game. In the following, we first explain why they cannot be applied to our game. Then we design a payoff allocation which is proved to be in the core of the game.

5.1 Conventional Payoff Allocations

5.1.1 Shapley Value

In the coalitional game theory, the Shapley value is a well-developed allocation, which can be calculated in a closed form. Specifically, it can be calculated as

$$\pi_i = \sum_{\mathcal{S}\subseteq\mathcal{N}\setminus\{i\}} \frac{|\mathcal{S}|!(N-|\mathcal{S}|-1)!}{N!} [v(\mathcal{S}\cup\{i\}) - v(\mathcal{S})], \quad (20)$$

where $|\mathcal{S}|$ is the size of coalition $\mathcal{S}$. Furthermore, four properties, i.e., *efficiency*, *symmetry*, *dummy player* and *additivity*, can be satisfied simultaneously using the Shapley value. Unfortunately, the Shapley value is not guaranteed to be in the core of the coalitional game. Although it can be guaranteed to be in the core of a convex coalitional game defined as

$$v(\mathcal{S}\cup\mathcal{S}') \geq v(\mathcal{S}) + v(\mathcal{S}') - v(\mathcal{S}\cap\mathcal{S}'),$$

where $\mathcal{S}$ and $\mathcal{S}'$ are any two sub-coalitions of $\mathcal{N}$, our game is not convex as shown in Appendix B, which implies that the Shapley value is not guaranteed to be in the core of our game. Therefore, the Shapley value cannot be used in our game.

5.1.2 Nucleolus

In the coalitional game theory, the *excess* defined as

$$e(\boldsymbol{\pi}, \mathcal{S}) = v(\mathcal{S}) - \sum_{i\in\mathcal{S}} \pi_i \quad (21)$$

is used to model the dissatisfaction of data centers in coalition $\mathcal{S}$ for an allocation vector $\boldsymbol{\pi}$. Different from the Shapley value which satisfies the four properties mentioned before, the nucleolus focuses on designing an allocation such that the dissatisfaction of data centers can be minimized. The nucleolus of a coalitional game always exists and is unique. Moreover, if the core is nonempty, the nucleolus lies in the core. However, calculating the nucleolus is computationally expensive, i.e., $O(2^N)$ linear programs need to be solved. Therefore, the nucleolus is not efficient to be used for our problem.

5.2 Proposed Allocation

Given that the traditional payoff allocations are not applicable in our game, in the following, we propose an alternate solution, which can stabilize the grand coalition.

According to Lemma 1, the data centers will form into the grand coalition $\mathcal{N}$ and sign the forward contract based on the fully aggregated DR capacity $\sum_{i=1}^{N} D_i(\lambda_i)$ so that the maximum total expected profit can be obtained.

To divide the expected profit among the data centers, we calculate the expected payoff for each tenant i as follows:

$$\pi_i = \frac{\partial v_{\boldsymbol{\alpha}}(\mathcal{N})}{\partial \alpha_i}\Big|_{\boldsymbol{\alpha}=\mathbf{1}}, \; \forall i \in \mathcal{N}. \quad (22)$$

In the following, we show that the payoff allocation (22) has four properties denoted as *stability*, *fairness*, *low computation complexity* and *efficiency*.

First of all, we discuss the stability property. The stability property is two fold. Firstly, the proposed payoff allocation guarantees that no deviation is profitable to any sub-coalition of $\mathcal{N}$. Secondly, no data center has the ability to manipulate its expected profit, i.e., no data center can improve its expected profit by forming into small coalitions first and participate into the CBP. In the following, we first show that our proposed payoff allocation (22) guarantees no sub-coalition can benefit from deviation using the following theorem.

THEOREM 2. *Our coalitional game has a nonempty core. Furthermore, the payoff allocation* (22) *lies in the core of our coalitional game.*

PROOF. See Appendix C for the detailed proof. □

Note that monetary payoff is adopted in our coalitional game, which is transferrable among data centers. According to the Bondareva-Shapley theorem [28], our game is balanced since it has a nonempty core. According to the definition (19), Theorem 2 also shows that if the expected profit generated by the grand coalition is allocated using (22), no sub-coalition has the incentive to deviate from it. Next, we discuss how the proposed payoff allocation prevents the data centers manipulating their expected profit. Since cooperation is encouraged under our setting, the data centers might potentially manipulate the game when forming into groups, i.e., the group may have the capability to improve its payoff by behaving strategically when participating into CBP. For example, the data centers with high DR capacities can form into a group, whose aggregated DR capacity occupies a large fraction of the fully aggregated DR capacity of all data centers. The group may reserve part of its DR capacity when participating into CBP, expecting its expected profit can be improved. The proposed payoff allocation prohibited the gain of market power in our coalitional game. Since the payoff received by each data center is the derivative with respect to its own aggregation level, participating into the DR program by forming into a group beforehand cannot increase the payoff received by each data center in the group comparing to participating individually, i.e., following the payoff allocation (22), no coalition has the ability to manipulate its expected profit.

One desired property of the Shapley value is fairness. Our proposed allocation also satisfies fairness in the following sense. For each data center i, its expected payoff π_i captures its marginal contribution to the aggregation. Therefore, two data centers with the same contribution should receive the same expected payoff. Since our payoff allocation not only stabilizes the coalitional game, but also satisfies fairness, it outperforms the Shapley value in our case.

Comparing with the nucleolus, the payoff allocation (22) has its strength so that it is more suitable for our coalitional game. According to (22), our proposed payoff allocation only requires to solve $O(N)$ equations, while the nucleolus requires to solve $O(2^N)$ linear programs. It is obvious that our proposed payoff allocation has much lower computation complexity than the nucleolus. Therefore, comparing to the nucleolus, the proposed allocation is more scalable to the scenario with a large set of data centers participating into DR program.

Table 1: Simulation Parameters

	$T_i^{\max}$ (ms)	M_i	μ_i(request/s)
data center 1	6	20000	500
data center 2	5	18000	450
data center 3	4	16000	350
data center 4	3	14000	300

The last observation is that the proposed allocation is efficient. Based on our previous analysis, no coalition has the ability to generate higher expected profit than the grand coalition with fully aggregated DR capacity. Following the payoff allocation (22), the total expected profit generated by the grand coalition is distributed to all data centers based on their marginal contributions, i.e., no profit loss is incurred during the payoff allocation process.

6. NUMERICAL EVALUATION

In this section, trace driven simulations are conducted. We first introduce our simulation settings. Then we show the simulation results to illustrate the effectiveness of our proposed approach.

6.1 Simulation Setup

Data center setup. We consider a set of four independent data centers $\mathcal{N} = \{1, 2, 3, 4\}$. The total number of servers for each data center is 20000, 18000, 16000 and 14000, respectively. We assume the idle power for all servers is 50 W. Furthermore, the PUE for all data centers are set as 1.5.

Workload description. The real-world workload data we use for evaluation is the cluster trace data of Google [31, 41]. The original workload is recorded for one month. We repeat it and extend it to 42 days, i.e. the length of the workload trace is 1000 hours.

The service rate of a server from each data center is set to be 500, 450, 350, and 300 requests per second, respectively. The maximum delay tolerance of the workload of each data center is set to be 6, 5, 4 and 3 ms, respectively. All the parameter settings are summarized in Table 1.

CBP description. In our simulations, we assume the CBP requires the data centers to sign the forward contracts day ahead. Furthermore, the CBP signals are issued hourly, i.e. the data centers are required to sign 24 forward contracts with the utility company day ahead. We fix the price p for the forward contract as \$1 per kW, the penalty rate a as \$1.5 per kW, and the reward rate b as \$0.5 per kW.

Finally, all our simulations are conducted on a desktop with an Intel i7 3.4 GHz CPU and 8GB RAM using MATLAB R2015a.

6.2 Performance Evaluation

In this part, we present the simulation results. We simulate how the data centers participate into the DR program under the cooperative setting. To demonstrate the benefits brought by cooperation, the individual participation without cooperation scenario is chosen as the baseline for comparison. In the following, all numerical results are averaged across days.

First, we construct the empirical distribution of the normalized workload arrivals as shown in Fig. 1. Then the empirical distribution of the DR capacity of each data center
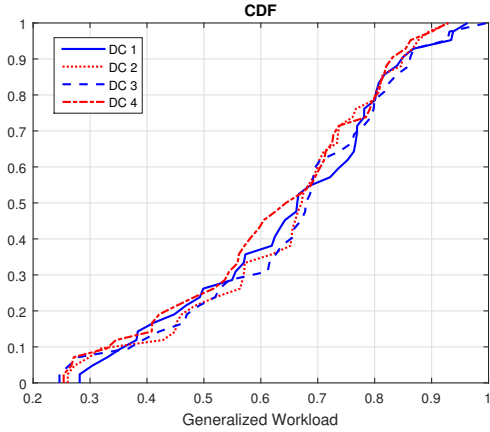
Figure 1: The empirical distribution of the normalized workload arrivals of the data centers (DC) at hour 10.
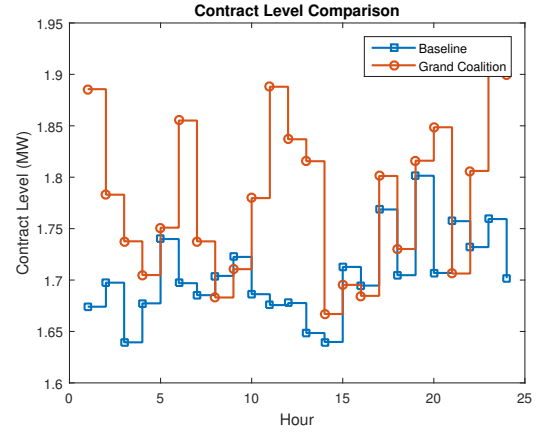

Figure 2: The empirical distribution of the DR capacities of the data centers (DC) at hour 10.

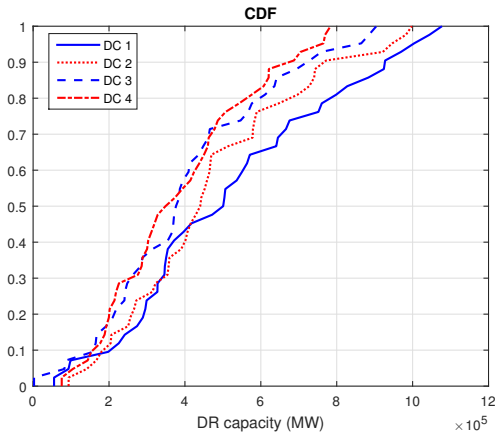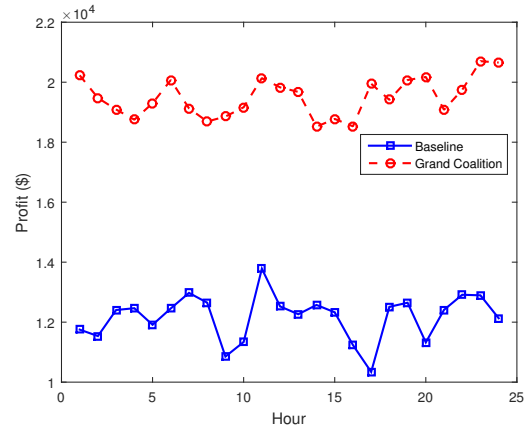Figure 3: The contract level of the baseline scenario and the grand coalition scenario for one day.

Figure 4: The total profit of the baseline scenario and the grand coalition scenario along a 24 hour time span. The profit for the baseline scenario is obtained by summing up the contracts of each data center.

i at each hour h obtained using (7) can be constructed to approximate the theoretical distribution at the same hour. Fig. 2 presents the CDF of the DR capacity of each data center at hour 10.

According to Theorem 1, we can obtain the optimal contract $C_{\mathcal{S}}^*$ for any coalition $\mathcal{S}$. Since the data centers are profit maximizing, given the DR capacity is uncertain due to the random workload arrivals, each data center will participate into the DR program conservatively. However, under the cooperative setting, the coalition will behave more aggressively seeking larger profits. We use Fig. 3 to show the difference of the data center DR commitment in advance between the baseline scenario and the grand coalition scenario during one day, which can be used to characterized how conservative are the data centers when participate into the CBP. The contract levels shown in Fig. 3 are averaged across days. In Fig. 3, the contract level for the baseline scenario is obtained by summing up the contracts of all data centers. It can be observed that the grand coalition commits more amount of DR than the summation of the commitment of each individual data center without cooperation. The reason is that after cooperation, the grand coalition has higher confidence on the aggregated DR capacity. Therefore, it is

to the joint benefit of all data centers in the grand coalition to behave less conservatively so that the maximum profit can be obtained. The improvement of the confidence on the aggregated DR capacity comes from the statistical diversity of the workload arrival of different data center.

In Fig. 4 we further show that the less conservative behavior of the grand coalition can lead to a larger profit than the baseline scenario. Two observations can be made in Fig. 4. The first observation is that the profit generated by grand coalition is much higher than the summation of the profit generated individually. The average improvement is 60.75%. This validates the supper additivity property presented in Section 4. Another observation is that in Fig. 4, the curve for the total profit of the grand coalition is smoother than that of the curve for the baseline scenario, i.e., the variance of the profit generated by grand coalition is much smaller than that of the baseline scenario. Specifically, the variance of the total profit of grand coalition is 35.94% lower than that of the baseline scenario. This observation indicates that cooperation can stabilize the profit generation ability of the coalition. Both observations demonstrate that the data cen-
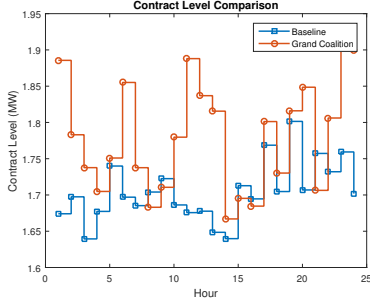
241
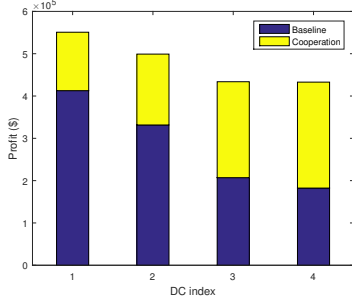
(a) Profit of each data center at hour 10 when price penalty ratio $r = 2/3$.
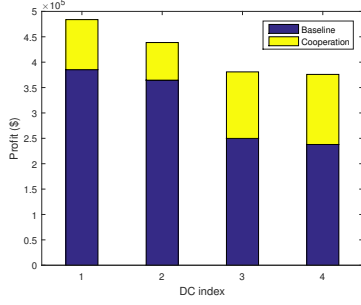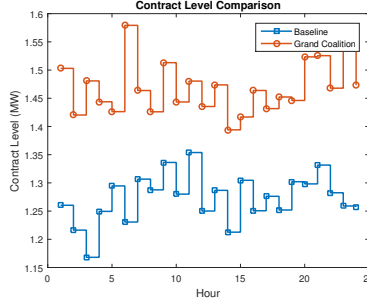

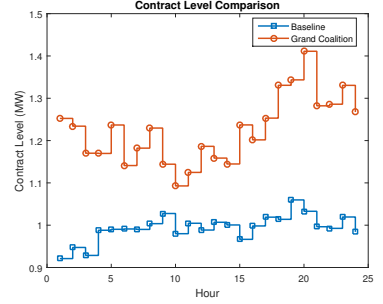(b) Profit of each data center at hour 10 when price penalty ratio $r = 1/2$.
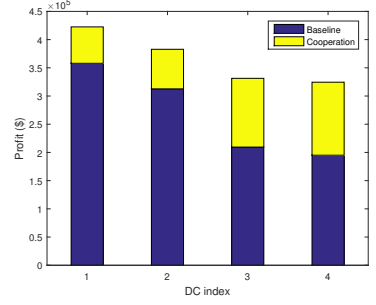
(c) Profit of each data center at hour 10 when price penalty ratio $r = 1/3$.

(d) Contract level over one day time period when price penalty ratio $r = 2/3$.

(e) Contract level over one day time period when price penalty ratio $r = 1/2$.

(f) Contract level over one day time period when price penalty ratio $r = 1/3$.

Figure 5: Fig. 5a to Fig. 5c show how the price penalty ratio impacts the expected profit of each data center. Fig. 5d to Fig. 5f show how the grand coalition behaves in the forward market.
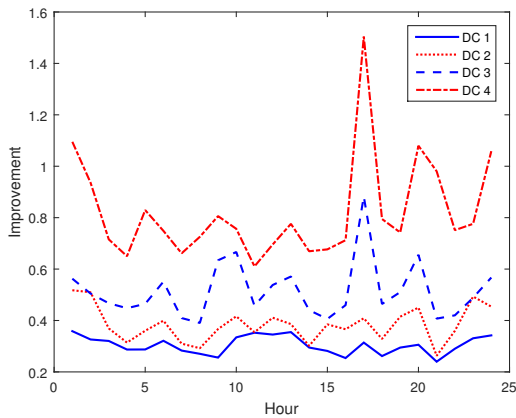

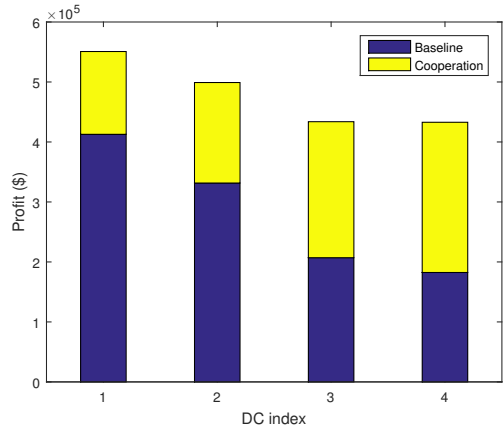Figure 6: The percentage improvement of each data center after cooperation.

Figure 7: The profit of each data center at hour 10.

ters will benefit from the cooperation.

Next, we show how much each data center can benefit from the cooperation process. Fig. 6 quantifies the percentage improvement of each data center after cooperation. We can observe that data center 1 benefits least from cooperation. On the contrary, data center 4 benefits most from the cooperation. The reason is two fold. First, the expected DR capacity of data center 1 is highest, while that of data center 4 is lowest. Therefore, the expected profit of data center 1 is much higher than that of data center 4 under the baseline scenario, which lowers the improvement of data center

1. Second, according to Fig. 2, we can see that comparing to data center 4, data center 1 has higher flexibility and lower probability being penalized. Therefore, data center 4 can benefit more from cooperation by exploiting the statistical diversity. To further clarify how the data centers benefit from cooperation, in the following, we show the profit obtained by each data center in a specific hour. Fig. 7 shows each data center's profit at hour 10. The height of each bar represents the profit obtained by each data center at hour 10. The yellow bar on the top of each bar is the profit improvement obtained via cooperation. It can be observed that data center 1 obtains the largest profit under both scenarios,

Table 2: The profit improvement of each data center versus the price penalty ratio.

	$r = 2/3$	$r = 1/2$	$r = 1/3$
data center 1	33.44%	25.58%	18.03%
data center 2	41.63%	27.03%	20.64%
data center 3	66.64%	39.45%	30.81%
data center 4	75.66%	41.43%	33.09%

while data center 4 obtains the lowest profit. Furthermore, we can see that data center 4 benefits most from cooperation. This is because cooperation decreases the uncertainty level of its DR capacity.

Finally, we analyze how the profit received by each data center will change with respect to the parameters defined by the CBP, i.e. how the forward price p, the penalty rate a and the reward rate b will impact the expected profit obtained by each data center. In the following we explore the impact by adjusting the price penalty ratio r defined as

$$r = \frac{p}{a}. \tag{23}$$

In our simulations, we fix the forward contract price p and the reward rate b, and set the penalty rate a as \$1.5 per kW, \$2 per kW and \$3 per kW to adjust the price penalty ratio r. Fig. 5 shows the profit obtained by each data center at hour 10 with different price penalty ratio r. The first observation is that with the decrease of the price penalty ratio, the expected profit of each data center decreases. The reason is two fold. The most intuitive explanation is that increasing the penalty rate will surely decrease the expected profit if the DR capacity distribution is fixed. Furthermore, note that the price penalty ratio indicates how heavily the CBP will penalize the data center with negative imbalance. Therefore, decreasing the price penalty ratio will lead the data centers to behave more conservatively when participating into the CBP, which directly decreases the profit. The conservative level can be measured by the contract level as shown in Fig. 5d to Fig. 5f. It is obvious that the average forward contract level decreases with respect to the decrease of the price penalty ratio r. Specifically, the average contract levels of the grand coalition with price penalty ratio 2/3, 1/2 and 1/3 are 1.78 MW, 1.47 MW and 1.22 MW, respectively. Another observation is that decreasing the price penalty ratio will decrease the improvement obtained from cooperation, i.e, the fraction of the yellow bars decrease as the price penalty ratio decreases. The detailed result is listed in Table 2. The idea behind this observation is that the high penalty rate will result in conservative behaviors in case of possible heavy penalty when participating into the CBP. Without loss of generality, in the following we consider an extreme case. When the penalty rate goes to infinity, the improvement from cooperation cannot increase since no matter how much the grand coalition commits in the forward contract, the penalty term will result in negative infinity profit.

7. CONCLUSION

Data centers are promising resources for DR programs. However, the randomness of data center DR capacities makes data centers unreliable DR resources. In this paper, we focus on how to make data centers reliable DR resources by considering data centers participating into CBP in which data centers are required to sign forward contracts for DR committment. We have shown that the actions in advance are unfavorable may drive the data centers out of the DR market if the uncertainties of the DR capacities in real time is not properly addressed. We have proposed that the data centers can cooperate with others and participate into CBP as a single DR provider to maximize expected profit by exploiting the statistical diversity. A coalitional game based approach has been used to study the cooperation process among the data centers, and we have shown that the grand coalition will be formed. To properly allocate the expected profit generated by coalition, we have designed a payoff allocation which can stabilize our game, satisfy fairness, and is efficient. Finally, trace driven simulation results are presented for demonstration.

8. REFERENCES

[1] Federal Energy Regulatory Commision. http://www.ferc.gov/legal/staff-reports/2015/demand-response.pdf.

[2] ISO New England Demand Response Working Group. http://www.iso-ne.com/.

[3] National Institute of Standards and Technology. http://www.nist.gov/public_affairs/releases/upload/smartgrid_interoperability_final.pdf.

[4] Pacific Gas and Electric Company. http://www.pge.com/en/mybusiness/save/energymanagement/cbp/index.page.

[5] Southern California Edison. https://www.sce.com/wps/wcm/connect/06f06444-92dd-4ca4-a32d-2454f09fb321/SCE_Capacity+Bidding+Program_NR-2232-V1-0513.pdf?MOD=AJPERES.

[6] United States Department of Energy. http://energy.gov/sites/prod/files/oeprod/DocumentsandMedia/DOE_Benefits_of_Demand_Response_in_Electricity_Markets_and_Recommendations_for_Achieving_Them_Report_to_Congress.pdf.

[7] M. H. Albadi and E. El-Saadany. A summary of demand response in electricity markets. In *IEEE Power Engineering Society General Meeting*, volume 78, pages 1989–1996, 2008.

[8] R. Brown et al. Report to congress on server and data center energy efficiency: Public law 109-431. Technical report, Lawrence Berkeley National laboratory, 2008.

[9] L. Chen, N. Li, S. H. Low, and J. C. Doyle. Two market models for demand response in power networks. In *Proc. IEEE Int. Conf. on Smart Grid Communications (SmartGridComm)*, volume 10, pages 397–402, 2010.

[10] N. Chen, X. Ren, S. Ren, and A. Wierman. Greening multi-tenant data center demand response. In *ACM SIGMETRICS Performance Evaluation Review*, 2015.

[11] R. A. Collins. The economics of electricity hedging and a proposed modification for the futures contract for electricity. *Power Systems, IEEE Transactions on*, 17(1):100–107, 2002.

[12] M. Ghamkhari and H. Mohsenian-Rad. Data centers to offer ancillary services. In *Proc. IEEE Int. Conf. on Smart Grid Communications (SmartGridComm)*, pages 436–441. IEEE, 2012.

[13] G. Ghatikar. Demand response opportunities and enabling technologies for data centers: Findings from field studies. Technical report, PG&E/SDG&E/CEC/LBNL, 2012.

[14] Y. Guo, Z. Ding, Y. Fang, and D. Wu. Cutting down electricity cost in internet data centers by using energy storage. In *Global Telecommunications Conference (GLOBECOM)*, pages 1–5, 2011.

[15] Y. Guo, Y. Gong, Y. Fang, P. P. Khargonekar, and X. Geng. Energy and network aware workload management for sustainable data centers with thermal storage. *Parallel and Distributed Systems, IEEE Transactions on*, 25(8):2030–2042, 2014.

[16] C. Lefurgy, X. Wang, and M. Ware. Server-level power control. In *Proc. Int. Conf. on Autonomic Computing (ICAC)*, pages 4–4. IEEE, 2007.

[17] N. Li, L. Chen, and S. H. Low. Optimal demand response based on utility maximization in power networks. In *Power and Energy Society General Meeting*, pages 1–8. IEEE, 2011.

[18] M. Lin, A. Wierman, L. L. Andrew, and E. Thereska. Dynamic right-sizing for power-proportional data centers. *IEEE/ACM Transactions on Networking (TON)*, 21(5):1378–1391, 2013.

[19] Z. Liu, Y. Chen, C. Bash, A. Wierman, D. Gmach, Z. Wang, M. Marwah, and C. Hyser. Renewable and cooling aware workload management for sustainable data centers. In *ACM SIGMETRICS Performance Evaluation Review*, volume 40, pages 175–186, 2012.

[20] Z. Liu, M. Lin, A. Wierman, S. H. Low, and L. L. Andrew. Greening geographical load balancing. In *Proc. of the ACM SIGMETRICS joint Int.Conf. on Measurement and modeling of computer systems*, pages 233–244, 2011.

[21] Z. Liu, I. Liu, S. Low, and A. Wierman. Pricing data center demand response. *SIGMETRICS Perform. Eval. Rev.*, 42(1):111–123, June 2014.

[22] Z. Liu, A. Wierman, Y. Chen, B. Razon, and N. Chen. Data center demand response: Avoiding the coincident peak via workload shifting and local generation. *Performance Evaluation*, 70(10):770–791, 2013.

[23] D. C. map. http://www.datacentermap.com/.

[24] A.-H. Mohsenian-Rad and A. Leon-Garcia. Coordination of cloud computing and smart power grids. In *Proc. IEEE Int. Conf. on Smart Grid Communications (SmartGridComm)*, pages 368–372. IEEE, 2010.

[25] A.-H. Mohsenian-Rad and A. Leon-Garcia. Distributed internet-based load altering attacks against smart power grids. *Smart Grid, IEEE Transactions on*, 2(4):667–674, 2011.

[26] R. B. Myerson. *Game theory: analysis of conflict*. Harvard University Press, 1991.

[27] NRDC. Data Center Efficiency Assessment. https://www.nrdc.org/energy/files/data-center-efficiency-assessment-IP.pdf.

[28] M. J. Osborne and A. Rubinstein. *A course in game theory*. MIT press, 1994.

[29] G. Owen. *Game Theory, 3rd edition*. UK: Academic Press, 1995.

[30] L. Rao, X. Liu, L. Xie, and W. Liu. Minimizing

[31] electricity cost: optimization of distributed internet data centers in a multi-electricity-market environment. In *Proc. IEEE Conf. on Computer Communications (INFOCOM)*, pages 1–9. IEEE, 2010.

[31] C. Reiss, J. Wilkes, and J. L. Hellerstein. Google cluster-usage traces: format + schema. Technical report, Google Inc., Mountain View, CA, USA, Nov. 2011. Revised 2012.03.20. Posted at http://code.google.com/p/googleclusterdata/wiki/TraceVersion2.

[32] A. E. Roth. *The Shapley value: essays in honor of Lloyd S. Shapley*. Cambridge University Press, 1988.

[33] F. Schneider, D. Klabjan, and U. W. Thonemann. Incorporating demand response with load shifting into stochastic unit commitment. *Social Science Research Network (SSRN)*, 2013.

[34] L. S. Shapley. A value for n-person games. *Operations Research*, 1952.

[35] E. Tanlapco, J. Lawarrée, and C.-C. Liu. Hedging with futures contracts in a deregulated electricity industry. *IEEE Power Engineering Review*, 7(22):54, 2002.

[36] N. H. Tran, S. Ren, Z. Han, S. M. Jang, S. I. Moon, and C. S. Hong. Demand response of data centers: A real-time pricing game between utilities in smart grid. In *USENIX Int Worksh. on Feedback Computing*, 2014.

[37] R. Urgaonkar, B. Urgaonkar, M. J. Neely, and A. Sivasubramaniam. Optimal power cost management using stored energy in data centers. In *Proc. ACM SIGMETRICS joint Int Conf. on Measurement and Modeling of computer systems*, pages 221–232, 2011.

[38] K. Wang, M. Lin, F. Ciucu, A. Wierman, and C. Lin. Characterizing the impact of the workload on the value of dynamic resizing in data centers. In *ACM SIGMETRICS Performance Evaluation Review*, volume 40, pages 405–406. ACM, 2012.

[39] A. Wierman, L. L. Andrew, and A. Tang. Power-aware speed scaling in processor sharing systems. In *Proc. IEEE Conf. on Computer Communications (INFOCOM)*, pages 2007–2015. IEEE, 2009.

[40] A. Wierman, Z. Liu, I. Liu, and H. Mohsenian-Rad. Opportunities and challenges for data center demand response. In *Green Computing Conference (IGCC), 2014 International*, pages 1–10, 2014.

[41] J. Wilkes. More Google cluster data. Google research blog, Nov. 2011. Posted at http://googleresearch.blogspot.com/2011/11/more-google-cluster-data.html.

[42] H. Xu and B. Li. Reducing electricity demand charge for data centers with partial execution. In *Proc. of Int. Conf. on Future Energy Systems*, pages 51–61. ACM, 2014.

[43] Y. Yao, L. Huang, A. Sharma, L. Golubchik, and M. Neely. Data centers power reduction: A two time scale approach for delay tolerant workloads. In *Proc. IEEE Conf. on Computer Communications (INFOCOM)*, pages 1431–1439, 2012.

[44] L. Zhang, S. Ren, C. Wu, and Z. Li. A truthful incentive mechanism for emergency demand response in colocation data centers. In *Proc. IEEE Conf. on Computer Communications (INFOCOM)*, 2015.

[45] R. Zhou, Z. Li, and C. Wu. An online procurement auction for power demand response in storage-assisted smart grids. In *Proc. IEEE Conf. on Computer*

Communications (INFOCOM), 2015.

[46] Z. Zhou, F. Liu, Z. Li, and H. Jin. When smart grid meets geo-distributed cloud: An auction approach to datacenter demand response. In *Proc. IEEE Conf. on Computer Communications (INFOCOM)*, 2015.

APPENDIX

A. PROOF OF LEMMA 1

First we prove the positive homogeneity property. For any scalar $\varepsilon \geq 0$, we can obtain the CDF of the random variable $\varepsilon A_{\boldsymbol{\alpha},\mathcal{S}}(\boldsymbol{\lambda})$ as

$$F_{\varepsilon\boldsymbol{\alpha},\mathcal{S}}(z) = \Pr\left(\varepsilon A_{\boldsymbol{\alpha},\mathcal{S}}(\boldsymbol{\lambda}) \leq z\right) = F_{\mathcal{S}}\left(\frac{z}{\varepsilon}\right).$$

Let $F_{\varepsilon\boldsymbol{\alpha},\mathcal{S}}^{-1}(\tau)$ be the quantile function of $\varepsilon \sum_{i=1}^{S} \alpha_i D_i$. Then according to the definition of the quantile function, the quantile function $F_{\varepsilon\boldsymbol{\alpha},\mathcal{S}}^{-1}(\tau)$ can be characterized as

$$F_{\varepsilon\boldsymbol{\alpha},\mathcal{S}}^{-1}(\tau) = \varepsilon F_{\mathcal{S}}^{-1}(\tau).$$

According to (11), we can prove the positive homogeneity property as follows:

$$\begin{aligned}
&v_{\varepsilon\boldsymbol{\alpha}}(\mathcal{S}) \\
&= a \int_0^{\tau^*} F_{\varepsilon\boldsymbol{\alpha},\mathcal{S}}^{-1}(z)\,\mathrm{d}z - b \int_{\tau^*}^1 F_{\varepsilon\boldsymbol{\alpha},\mathcal{S}}^{-1}(z)\,\mathrm{d}z - \\
&= \varepsilon\left(a \int_0^{\tau^*} F_{\mathcal{S}}^{-1}(z)\,\mathrm{d}z - b \int_{\tau^*}^1 F_{\mathcal{S}}^{-1}(z)\,\mathrm{d}z - \right) \\
&= \varepsilon v_{\boldsymbol{\alpha}}(\mathcal{S}).
\end{aligned}$$

Next we prove the supper additivity property. Suppose the two optimal forward contracts corresponding to the fully aggregated DR capacities $\varepsilon \sum_{i=1}^{S_1} \alpha_i D_i$ and $\varepsilon \sum_{i=1}^{S_2} \alpha_i D_i$ are $C_{\mathcal{S}_1}^*$ and $C_{\mathcal{S}_2}^*$, respectively. We further denote the optimal solution for coalition $\mathcal{S}_1 \cup \mathcal{S}_2$ as $C_{\mathcal{S}_1 \cup \mathcal{S}_2}^*$. Note that $C_{\mathcal{S}_1}^* + C_{\mathcal{S}_2}^*$ is a feasible solution for the expected profit maximization problem of coalition $\mathcal{S}_1 \cup \mathcal{S}_2$. Then it is straightforward to see that $v(\mathcal{S}_1) + v(\mathcal{S}_2) \leq v(\mathcal{S}_1 \cup \mathcal{S}_2)$.

Third, we prove the nondecreasing property. Because of the supper additivity property, in the following, we focus on the grand coalition. Suppose we have two aggregation level vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}'$, where $\boldsymbol{\alpha}' \succeq \boldsymbol{\alpha}$. Then it is obvious that for any element in vector $\boldsymbol{\alpha}' - \boldsymbol{\alpha}$,

$$0 \leq \alpha' - \alpha \leq 1.$$

Using the supper additivity property, we have

$$v_{\boldsymbol{\alpha}'}(\mathcal{N}) \geq v_{\boldsymbol{\alpha}}(\mathcal{N}) + v_{\boldsymbol{\alpha}'-\boldsymbol{\alpha}}(\mathcal{N}). \tag{24}$$

The inequality above indicates the nondecreasing property.

B. PROOF OF NONCONVEX GAME

Here, we use a counterexample to prove our game is not convex. Consider a colo with a set of data centers $\mathcal{N} = \{1, 2, 3\}$ housing in it. Denote their normalized DR capacities as D_1, D_2 and D_3. Assume the distribution of the normalized DR capacity of any data center $i \in \mathcal{N}$ is

$$D_i = \begin{cases} 0.5 & \text{w.p. } 0.5 \\ 1 & \text{w.p. } 0.5 \end{cases} \quad \forall i \in \mathcal{N}. \tag{25}$$

We further assume that the realizations of D_3 always equal to that of D_2. Consider when the penalty rate a and reward rate b are deterministic. Suppose the forward contract price $p = 1.1$, $a = 1.5$, $b = 0.5$ and $e_1 = e_2 = e_3 = 0.1$. Then the characteristic function can be calculated as follows:

$$\begin{aligned}
v(\{1\}) = v(\{2\}) = v(\{3\}) &= 0.5; \\
v(\{1, 2\}) &= 1.375 \\
v(\{1, 3\}) &= 1.375; \\
v(\{2, 3\}) &= 1; \\
v(\{1, 2, 3\}) &= 2;
\end{aligned}$$

Consider two coalitions $\mathcal{S} = \{1, 2\}$ and $\mathcal{S}' = \{1, 3\}$. Then we have

$$v(\mathcal{S} \cup \mathcal{S}') = v(\mathcal{N}) = 2 < v(\mathcal{S}) + v(\mathcal{S}') - v(\{1\}) = 2.25.$$

Therefore, our coalitional game is not convex.

C. PROOF OF THEOREM 2

We prove Theorem 2 by demonstrating the payoff allocation (22) lies in the core of our coalitional game. Note that there are two requirements in the definition of the core (19). First we prove that our solution satisfies $\sum_{i \in \mathcal{N}} \pi_i = v(\mathcal{N})$ as defined in (19). Note that in our case the colo operator is trustworthy, and it will allocate all the profits to the data centers. Therefore, there is no welfare loss in our case. Then it is straightforward to see π satisfies $\sum_{i \in \mathcal{N}} \pi_i = v_{\mathbf{1}}(\mathcal{N})$. Next, we prove that our proposed payoff allocation satisfies $\sum_{i \in \mathcal{N}} \pi_i \geq v(\mathcal{S})$, $\forall \mathcal{S} \subseteq \mathcal{N}$ as defined in (19). Consider an arbitrary coalition $\mathcal{S}$, $\forall \mathcal{S} \subseteq \mathcal{N}$. If each data center $i \in \mathcal{S}$ holds $\omega_i = D_i / \sum_{i=1}^{N} D_i$ fraction of the aggregated DR capacity, then using positive homogeneity property, we have $\sum_{i \in \mathcal{S}} \pi_i = v_{\boldsymbol{\omega}}(\mathcal{S})$, where $\boldsymbol{\omega} = [\omega_1, \omega_2, \cdots, \omega_S]$. Since the grand coalition $\mathcal{N}$ explores more statistical diversity than coalition $\mathcal{S} \subseteq \mathcal{N}$, and $\boldsymbol{\omega}$ has no impact on the statistical diversity, the inequality above holds. Therefore, the second requirement holds for our proposed solution as well. Then it is straightforward to see that the payoff allocation (22) lies in the core of our coalitional game.