Fair Rewarding in Colocation Data Centers: Truthful Mechanism for Emergency Demand Response

Qihang Sun†, Chuan Wu†, Shaolei Ren‡ and Zongpeng Li§
†Department of Computer Science, The University of Hong Kong, Email: {qhsun,cwu}@cs.hku.hk
‡School of Computing and Information Sciences, Florida International University, Email: sren@fiu.edu
§Department of Computer Science, University of Calgary, Email: zongpeng@ucalgary.ca

Abstract—**Reducing servers' power usage in data centers upon utility's request has been emerging as a valuable demand response resource for enhancing power grid's efficiency and reliability, especially during emergency events (e.g., extreme weather) that result in electricity production shortage and put the grid in jeopardy. Nonetheless, for demand response in multi-tenant colocation data centers, operators may have to leverage expensive and environmentally-unfriendly diesel generation, because individual tenants manage their own servers' power usage without coordination and are typically charged by data center operators based on fixed power contracts that provide no incentives for demand response. This paper focuses on emergency demand response (EDR) and proposes an auction-based incentive mechanism, called FairDR, that incentivizes and coordinates tenants' energy reduction through financial rewards for enabling cost-effective and low-carbon EDR in colocation data center. FairDR decides tenants' energy reduction online without knowing a priori the future energy reduction requirements. It is proved that FairDR ensures tenants' truthfulness in the auction process, attains a bounded overall cost saving compared to the offline optimum which knows all the demands, and guarantees fairness (i.e., similar rewards are offered if tenants reduce the same amount of energy) that is largely absent in the existing auction mechanisms. Finally, trace-driven simulations are performed to validate our analysis and demonstrate that FairDR outperforms the existing mechanisms by improving fairness and achieving a good cost saving that is comparable to the offline optimum.**

I. Introduction

Large-scale data centers are power-hungry but highly-automated facilities, having a huge yet flexible power demand (*e.g.*, through deferrable workload shifting/migration). While the huge power appetite is unpleasant by itself, the great flexibility in data center's power consumption has been emerging as an ideal resource for demand response, especially for emergency demand response (EDR) where the power grid coordinates large electricity users for energy reduction in emergency situations (*e.g.*, continuous snow storms as recently in east U.S.) [1], [2].

While data center demand response has been investigated by numerous studies as well as validated through field tests (see [1], [3] and references therein), a vast majority of the existing efforts focus on owner-operated data centers (*e.g.*, Google data centers), where operators *fully* control the entire data center (including both servers and facility). By significant contrast, we investigate demand response in another distinct type of data center — colocation data center (or simply called "colocation") — where multiple tenants manage their own physical servers while the operator is only responsible for managing the facility, *e.g.* the cooling system and the power supply. Colocation is much less investigated than owner-operated data centers but very common in practice: by February 2015, there are more than 1,400 colocation data centers in the U.S. [4]; furthermore, most of the large data centers are colocations [5]. More importantly, unlike mega-scale owner-operated data centers, many colocation data centers are located in populated areas, such as Silicon Valley and New York city [4], where demand response is even more important (due to high power demand) than in rural areas, especially in case of emergency.

Although critical for power grid's efficiency, reliability and sustainability, demand response faces a unique challenge in colocation data centers, since individual tenants manage their own servers without coordination and sign long-term contracts with the operator based on reserved power capacities at fixed rates [6] that provide no incentives for participating in demand response. Thus, unlike owner-operated data centers that can easily modulate servers' power usage (e.g., through CPU frequency control), the colocation operator may have to leverage expensive and/or environmentally-unfriendly diesel generation to reduce grid power usage for demand response [7].

This paper aims at enabling cost-effective and low-carbon demand response in colocation data centers through an efficient mechanism that incentivizes and coordinates tenants to voluntarily cut their servers' power usage. In particular, we focus on EDR, which serves as the last line of defense for protecting the grid against cascading blackouts and, as of 2013, has taken up 87% of all demand response capacities across the U.S. [8]. EDR is also quickly growing as extreme events (e.g., weather, demand spikes) become increasingly frequent. For example, in PJM, a primary regional transmission organization in the U.S., the total capacity of EDR energy reduction increases from 1,700MW to 10,800MW during 2006-2011 [9]. When EDR is triggered during an emergency event, multiple random energy reduction signals may arrive from the power grid over time (often hourly), depending on how long the emergency situation lasts. For example, on January 7, PJM's power grid issued energy reduction signals 11 times due to extremely low temperatures; on September 10, 2013, unusually hot weather triggered 4 energy reduction signals within PJM's service area [10]. Nonetheless, as shown later, unknown and random energy reduction signals arriving online create significant technical challenges for our mechanism design.

This paper designs an auction mechanism, called *FairDR*, to respond to the entire sequence of energy reduction signals for EDR in colocation data centers. A salient feature of FairDR

This work was supported in part by a grant from Hong Kong RGC under the contract HKU 718513, U.S. NSF CNS-1423137, and CNS-1453491 (CAREER).

978-1-4673-7113-1/15/$31.00 2015 IEEE

is that the bidding information (e.g., tenants' willingness to cut power) is collected only once from the tenants, while tenants' actual energy reduction is decided online upon the arrival of each EDR signal, without knowing any future signals. Our specific contributions are summarized as follows:

- ▷ We formulate and investigate the problem of energy reduction allocation among tenants during EDR in a colocation data center, with the goal of maximizing social cost saving. The innovation of the model lies in that: we consider a realistic EDR scenario where multiple energy reduction signals are often triggered over a time window (*e.g.* eight hours), and achieve truthfulness throughout the auction period (the *entire* EDR process) in an *online* manner without prior knowledge of future signals.
- ▷ We design FairDR based on an effective two-level randomization strategy in its tenant selection and reward pricing. Regardless of the energy reduction requirements, FairDR guarantees that truthful bids are always reported by tenants and similar rewards are offered to tenants with similar energy reduction (i.e., fairness). In addition, FairDR reduces colocation's cost for EDR as opposed to the diesel-only solution in an online manner, its online decisions for allocating energy reduction to tenants attain an expected competitive ratio of ≈ 3.2 under realistic settings in terms of overall social cost saving, as compared to the offline optimum in the entire EDR span. To the best of our knowledge, this represents the first efficient mechanism to handle online EDR signal arrivals with long-term performance guarantees.
- ▷ We compare FairDR with the existing best-known algorithm for colocation EDR, i.e., Truth-DR [11]. We show that FairDR achieves almost perfect fairness in rewarding yet incurs little cost increase compared to Truth-DR.

The rest of the paper is organized as follows. We present the problem model in Sec. II. In Sec. III, we identify the design challenges, develop the detailed auction mechanism, and provide theoretical analysis of its properties. Simulation results are presented in Sec. IV. We discuss related work and conclude the paper in Sec. V and Sec. VI.

II. Preliminaries and Problem Formulation

A. Problem Model

We consider a colocation data center with N tenants, operated by a colocation operator. The tenants rent space and power from the operator, to house their own servers. Suppose the EDR lasts for an overall duration of T time slots (which can also be the maximum duration that the colocation data center has agreed to participate as stipulated by contracts). Multiple (random) reduction signals arrive online throughout $[1, \dots, T]$. Let s_t denote the amount of energy reduction required at $t \in [1, \dots, T]$. We use $s_t = 0$ to indicate that no reduction signal is received at t. Note that the data center typically sign contracts with grid operators and is required to fulfill the energy reduction signal for EDR, while non-compliance incurs heavy penalties [9].

To incentivize tenants' server energy reduction as a (possibly partial) substitute of diesel generation for EDR, the colocation operator needs to offer financial rewards to participating tenants. In our study, the colocation operator performs an auction to allocate the energy reduction $s_t > 0, \forall t = 1, \dots, T$, to tenants and pays monetary remuneration accordingly. Here, the operator is the auctioneer with online arrivals of "supply" (i.e., energy reduction requests), while the tenants are bidders.

We investigate a bidding model where the operator solicits bids from the (voluntarily participating) tenants only once through the whole course of the EDR, upon the start of the EDR. In particular, a bid from tenant $i \in \mathcal{N} = \{1, 2, \dots, N\}$ is a three-tuple (e_i, c_i, b_i): e_i denotes the agreed maximum overall energy reduction by tenant i throughout the entire course of the EDR, c_i is the maximal amount of energy that tenant i is willing to cut per time slot,[2] and b_i is the cost per unit energy reduction *claimed* by tenant i. We use v_i to represent the true unit cost of tenant i, and as shown later, our mechanism design will guarantee $b_i = v_i$.

Such a bidding model is reasonable as follows: (1) It is practically more difficult for a tenant to come up with multiple bids (if bids are solicited repeatedly upon arrival of each reduction signal), while deciding one bid for the entire EDR is more realistic, leading to easier adoption of the auction mechanism. (2) With a great flexibility in delay-tolerant workloads (e.g., by rescheduling job execution or routing some workload to other data centers), a tenant can readily modulate its energy usage and also estimate its overall energy reduction capability based on predicted workloads throughout the estimated duration of the emergency event. We also allow each tenant to set an upper bound on the energy reduction per time slot, in order to maintain the performance and/or continuity of its services, and such upper bound is naturally limited by the tenants' maximum energy usage per time slot. For example, a tenant's peak power usage is 200kW, its average power consumption for maintaining satisfactory service performance is 80% of the peak power, and the maximal power to cut per time slot is 45% of the peak power, in order to maintain a satisfactory performance level; given an 8-hour EDR event divided into 8 hourly time slots, the tenant can set its $e_i = 200 * (1 - 80\%) * 8 = 320$kWh and $c_i = 200 * 0.45 =$90kWh (per hour). (3) The true cost of the tenant can be estimated at the tenant's own discretion, *e.g.*, based on the wear-and-tear cost for turning servers on and off and/or the service performance degradation incurred.

Based on the collected bids, the operator decides, upon receiving each reduction signal $s_t > 0$ from the power grid, the tenant(s)' energy reduction x_i in the upcoming time slot, without exceeding the tenant's overall allowed energy reduction e_i and energy reduction limit c_i per time slot. The operator also computes the reduction reward p_i for each selected tenant i per unit of energy reduction. In the case that the tenants' aggregate energy reduction is not enough to fulfill the required amount s_t, the operator resorts to its *diesel generator* to cover the shortage, at the cost δ for producing a unit of energy [7], [12].

B. Target Properties

We aim to achieve the following properties through mechanism design.

(i) *Truthfulness*, which guarantees that bidding its true cost is the best strategy for each tenant, is a much desired property

[2]Reducing server energy also proportionally reduces non-IT energy (*e.g.*, for cooling), which is attributed to tenants for notational convenience.

for mechanism design. Let $u_i = (p_i - v_i) \times x_i$ denote the utility of tenant i, where p_i is the offered reward and v_i is the true cost per unit energy reduction. We define truthfulness in the following.

Definition 1. (**Truthfulness**) *Our auction is truthful if each tenant $i \in \mathcal{N}$ achieves the largest utility u_i throughout the EDR event by reporting its true cost for each unit of power reduction, i.e., $b_i = v_i$, regardless of the bids of other tenants.*

(ii) *Fairness in rewards*. Our mechanism seeks to reduce the variance in rewards per unit energy reduction offered to different tenants. Such fairness is desirable to avoid significant reward differentiation among tenants. Nonetheless, fairness is rarely ensured in existing mechanism designs, which typically determine each winner's reward based on other tenants' bids in order to guarantee truthfulness and, consequently, provide different prices to different bidders even though they ask for the same allocation [11], [13]. By contrast, our mechanism addresses the less-studied reward fairness throughout the EDR event, in addition to truthfulness and social cost that are more common in the existing literature.

(iii) *Social cost saving maximization*. Let R represent the total amount of required energy reduction throughout the EDR event, *i.e.*, $R = \sum_{t=1}^{T} s_t$. The overall cost of the colocation operator can be computed as the sum of the rewards to the tenants and the overall cost due to diesel generation to cover the shortage in energy reduction, *i.e.*, $\sum_{i \in \mathcal{N}} p_i x_i + \delta(R - \sum_{i \in \mathcal{N}} x_i)$. The net cost of a tenant i is the cost due to energy reduction minus the reward from the operator, $b_i x_i - p_i x_i$ (assuming $b_i = v_i$, which is guaranteed by our mechanism). Hence, the overall social cost in the colocation is the sum of the costs of the operator and tenants, $\sum_{i \in \mathcal{N}} b_i x_i + \delta(R - \sum_{i \in \mathcal{N}} x_i)$.

Without tenants' participation in EDR through server energy reduction, the social cost would be the cost in using diesel generators to completely fulfill all the required energy reduction, which is δR. Hence, the social cost saving is $\delta R - (\sum_{i \in \mathcal{N}} b_i x_i + \delta(R - \sum_{i \in \mathcal{N}} x_i)) = \sum_{i \in \mathcal{N}} (\delta - b_i) x_i$. Our mechanism design aims to maximize this social cost saving, which facilitates the analysis and is essentially equivalent to minimizing the social cost.

III. Demand Response Auction with Online Reduction Signal Arrivals

We now present our auction algorithm for energy reduction allocation and tenant rewarding, as well as theoretical analysis on its properties.

A. Auction Algorithm

Before an EDR event is anticipated to occur, our auction algorithm chooses some tenants as *eligible tenants*, *i.e.*, those which will receive an opportunity to reduce energy consumption and get rewarded. The algorithm also decides the reward to be offered to each selected eligible tenant. Then, as the EDR event proceeds, whenever an energy reduction signal arrives, the algorithm selects the actual tenant(s) among the eligible tenants and decides the amount of energy reduction each selected tenant should cut. The details of the algorithm are described in the following.

TABLE I. Notation

N	# of tenants
$\mathcal{N}$	the set of tenants
e_i	maximally agreed, overall amount of energy reduction by tenant i
b_i	claimed cost per unit energy reduction at tenant i
v_i	true unit cost of tenant i
x_i	actual amount of energy reduction by tenant i
p_i	per-unit reward to tenant i
u_i	total utility of tenant i
S	sorted sequence of claimed per-unit costs in nondecreasing order
Δb_{max}	maximal difference between two adjacent per-unit bids in S
e_{min}	minimum among overall energy reduction capacities of all tenants
U	maximal ratio of overall energy reduction capacities of any two tenants
δ	per-unit cost using diesel generators
R	total amount of energy reduction from all reduction signals
m	sum of all tenants' overall energy reduction capacities
$\mathcal{E}$	the set of eligible tenants
$\mathcal{A}$	the random permutation of $\mathcal{E}$
c_i	maximal amount of energy to cut per time slot at tenant i
s_t	amount of energy reduction requested in signal at t
α	the minimal ratio of c_i to e_i, $\forall i \in \mathcal{N}$

Eligible tenant initialization

Let m denote the sum of maximum amounts of energy reduction of all tenants through the entire course of the EDR, *i.e.*, $m = \sum_{i=1}^{N} e_i$. The eligible tenant selection steps proceed as follows:

1) Pick q among the set of numbers $\{2^1, 2^2, \ldots, 2^i, \ldots, m\}$ uniformly at random, which represents the maximal, total amount of energy that the operator will ask tenants to cut through the entire course of the EDR.
2) Sort all tenants by the per-unit claimed cost in non-decreasing order.
3) From the start of the ordered sequence, find the minimal continuous sub-sequence where the sum of tenants' energy reduction capacity (e_i) is not less than q. The tenants in this sub-sequence, denoted by $\mathcal{E}$, are the eligible tenants.
4) Obtain a random permutation of the sequence $\mathcal{E}$, and denote the permutation by $\mathcal{A}$.

Intuitively, we should select tenants with low claimed costs (per unit energy reduction) as eligible tenants, to maximize social cost saving. In order to avoid tenants underclaiming their costs to get selected (truthfulness guarantee), we make the order of tenant selection independent of their claimed costs, using a randomization step (step 4 above). We should not randomize the entire sequence of N tenants in order to ensure that we are still mostly allocating power reduction to tenants with low costs (for cost saving guarantee), but only the subset of eligible tenants (picked on non-decreasing order of costs in steps 2 and 3), which may very likely be called on to reduce energy consumption during the EDR event. Such a randomization need explains why we need to pre-select the set of eligible tenants at the very beginning. To decide the set of eligible tenants, we need to 'guess' the total amount of energy reduction from all future signals that may arrive, q. By making a guess in this way (Step 1 above), the guessed

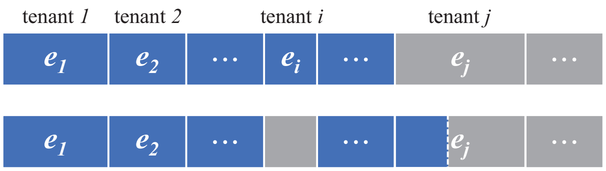

Fig. 1. An example on computing f_j and f_j^{-i}.

total amount of energy reduction is not less than half, nor more than twice of the actual amount. Our algorithm does not rely on any future information for this purpose, but only make a rough random guess (step 1) of q between 2 and m – overall possible energy cut from all tenants. With the proceeding of the EDR event, q may turn out to be larger than the actual total amount of energy reduction requests from the power grid ($\sum_{t=1}^{T} s_t$), in which case not all the eligible tenants in the permutation sequence $\mathcal{A}$ will be called on; q may also be possibly smaller than the actual total energy reduction request, in which case $q - \sum_{t=1}^{T} s_t$ will be covered by energy produced by the diesel generator. Nevertheless, we will be able to bound the ratio between the social cost saving achieved by our auction and the offline optimum, to be shown in Sec. III-B3.

Tenant rewarding

We compute the per-unit-energy-reduction reward for each eligible tenant at the beginning of the EDR event as

$$p_i = \frac{\sum_{j \in \{\mathcal{N} \setminus i\}} b_j f_j^{-i} - \sum_{j \in \{\mathcal{N} \setminus i\}} b_j f_j}{f_i} \tag{1}$$

where f_j denotes the maximally possible amount of energy reduction that a tenant j can be allocated, and f_j^{-i} is the maximally possible amount of energy reduction that a tenant j can be allocated if tenant i does not participate in the auction, both calculated according to q and e_i without considering per-time-slot energy reduction capacity c_i of each tenant i. Intuitively, the reward computed in (1) is a VCG price computed using f_is, the maximal possible amount of energy reduction which can be allocated. For an ineligible tenant j, we define $f_j = 0$. Fig. 1 illustrates the computation of f_j and f_j^{-i}, where each sequence of blocks represents the maximally allowed energy reduction amounts of all tenants (suppose tenants 1, 2, ..., are already sorted in non-decreasing order of claimed per-unit costs). Suppose $q = 8, e_i = 2, e_j = 3$, blocks in blue denote potential energy reduction during the EDR event and the corresponding tenants are eligible tenants. In the case that tenant i participates in the auction, tenant j is not eligible, and $f_i = 2, f_j = 0$; if tenant i is not a participant, tenant j would be eligible and $f_j^{-i} = 2$.

The reward p_j is set to the per-unit-energy-reduction externality of tenant i (in the ideal case without considering c_i's), as the overall cost difference of all tenants other than i when tenant i does not participate in the auction and when it participates, divided by the maximally possible amount of tenant i's energy reduction. Suppose the claimed unit cost of tenant j is $b_j = 5$ in the example in Fig. 1. Tenant i's total reward $p_i f_i$ equals the claimed cost for tenant j to reduce 2 units of energy consumption, *i.e.*, $b_j * 2 = 10$, and hence the per-unit reward of tenant i is $10/f_i = 5$.

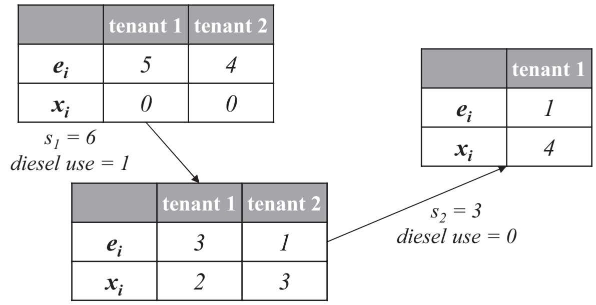

Fig. 2. An example on power reduction allocation: $c_1 = 2, c_2 = 3$.

We will show such a rewarding strategy guarantees both truthfulness and fairness in the auction, even if no per-time-slot energy reduction limits of tenants are involved. This per-unit reward for each eligible tenant is independent from the reduction signals, and can be easily computed at the beginning of the EDR event and immediately offered to tenants after they have completed their energy reduction upon each signal.

Winner determination and energy reduction allocation

Whenever an energy reduction signal arrives, the colocation operator chooses top-ranked tenants in the permutation sequence $\mathcal{A}$, maximally allocates the energy reduction amount $s_t > 0$ to the tenants according to their sequence in $\mathcal{A}$, as long as the overall energy reduction capacity e_i and per-time-slot energy reduction capacity c_i of a tenant are not exceeded. Fig. 2 illustrates the allocation process where $\mathcal{A}$ includes tenant 1 and tenant 2, and two reduction signals arrive at time slot 1 ($s_1 = 6$) and time slot 2 ($s_2 = 3$), respectively. We have $c_1 = 2$ and $c_2 = 3$ which are fixed, and $e_1 = 5$ and $e_2 = 4$ initially, which will be reduced when the respective tenant has reduced some energy consumption. Each table at the start (end) of an arrow shows e_i and x_i of the two tenants before (after) the arrival of the signal marked on the arrow. When s_1 arrives, tenant 1 is allocated two units of energy reduction due to the limitation of c_1, tenant 2 is allocated three units, and the remaining one unit is fulfilled by the diesel generator; when s_2 arrives, tenant 1 is allocated two units and tenant 2 one unit because there is only one unit left in e_2, and tenant 2 is removed from $\mathcal{A}$ after it has reduced energy consumption up to the initial e_2.

The complete auction algorithm, FairDR, is given in Alg. 1.

B. Theoretical analysis

1) Fairness in rewards: To show fairness of FairDR, we show the following theorem that bounds the difference in reward rates offered to different tenants, while our numerical results in Section IV will demonstrate that the actual difference is much smaller in practice.

Theorem 1. *With FairDR, the winning tenants receive similar per-unit rewards, and the difference between p_i and p_j of any two winners is upper bounded by $\frac{\lceil U \rceil + 1}{2} e_{min} \Delta b_{max}$, where $U = \frac{\max_{i \in \mathcal{N}} e_i}{\min_{j \in \mathcal{N}} e_j}$, $e_{min} = \min_{i \in \mathcal{N}} e_i$, and $\Delta b_{max} = \max_{i \in [1, N-1]} |b_i - b_{i+1}|$, with $\{b_1, b_2, \ldots, b_N\}$ being a sorted sequence of per-unit claimed costs of tenants in non-decreasing order.*

Algorithm 1: FairDR: Demand Response Auction with Online Reduction Signal Arrivals

input : $b_i, e_i, c_i, \forall i \in \mathcal{N}$
output: $x_i, p_i, \forall i \in \mathcal{N}$

Select q uniformly randomly from $\left\{2^1, 2^2, \ldots, 2^i, \ldots, \sum_{i=1}^{N} e_i\right\}$;
Sort all tenants in $\mathcal{N}$ in non-decreasing order of per-unit costs;
Choose tenants from the start of the ordered list to obtain the minimum subsequence $\mathcal{E}$, whose total energy reduction capacity is no smaller than q;
Permutate $\mathcal{E}$ randomly to get $\mathcal{A}$;
Compute p_i following Eq. 1, $\forall i \in \mathcal{N}$;
Initialize $x_i = 0, \forall i \in \mathcal{N}$;
for *arrival of each signal* s_t **do**
  if $\mathcal{A} \neq \varnothing$ **then**
    $i = FirstTenant(\mathcal{A})$; *%get the first tenant in* $\mathcal{A}$
    while $i \neq null$ *and* $s_t > 0$ **do**
      $d_{max} = \min\{e_i, c_i\}$;
      if $(d_{max} \geq s_t)$ **then**
        $e_i = e_i - s_t$;
        $x_i = x_i + s_t$;
        $s_t = 0$;
      else
        $e_i = e_i - d_{max}$;
        $x_i = x_i + d_{max}$;
        $s_t = s_t - d_{max}$;
        if $(e_i > 0)$ **then**
          $i = NextTenant(i, \mathcal{A})$; %get the next tenant following i in $\mathcal{A}$
        else
          $tenant\_to\_remove = i$;
          $i = NextTenant(i, \mathcal{A})$;
          $\mathcal{A} = \mathcal{A} \setminus tenant\_to\_remove$;
        end
      end
    end
  end
  if $s_t > 0$ **then**
    Use diesel generator to fulfil remaining energy reduction s_t;
  end
end

Here U represents the maximal ratio of overall reduction capacities of any two tenants, e_{min} is the minimal energy reduction capacity among all tenants, and Δb_{max} denotes the maximal difference of adjacent per-unit claimed costs in the sorted sequence.

Proof: Based on Eq. 1, we can derive the total reward of tenant i as

$$p_i f_i = \sum_{j \in \{\mathcal{N} \setminus i\}} b_j f_j^{-i} - \sum_{j \in \{\mathcal{N} \setminus i\}} b_j f_j \tag{2}$$

The random number q in general divides the overall reduction sequence as exemplified in Fig. 1, where tenants are ordered in non-decreasing order of their claimed per-unit costs, into two parts, denoted by the *eligible part* and the *ineligible part*, respectively.

Assume the claimed cost of the first unit in the ineligible part is G. For each eligible tenant, the lower bound of its per-unit reward is G and the upper bound of its per-unit reward is:

$$\begin{aligned} p_i \leq & G + \frac{1}{2} \cdot \frac{\left\lceil \frac{e_i}{e_{min}} \right\rceil \left(e_{min} \Delta b_{max} + \left\lceil \frac{e_i}{e_{min}} \right\rceil e_{min} \Delta b_{max} \right)}{\left\lceil \frac{e_i}{e_{min}} \right\rceil} \\ = & G + \frac{1}{2} e_{min} \Delta b_{max} \left(1 + \left\lceil \frac{e_i}{e_{min}} \right\rceil \right) \\ \leq & G + \frac{\lceil U \rceil + 1}{2} e_{min} \Delta b_{max} \end{aligned} \tag{3}$$

Hence, the maximal difference between per-unit rewards of any two tenants is $\frac{\lceil U \rceil + 1}{2} e_{min} \Delta b_{max}$. ■

2) Truthfulness: Before we show the truthfulness of FairDR, we first prove that it achieves individual rationality.

Lemma 1. *FairDR achieves individual rationality, i.e., no winning tenant's per-unit reward is less than its per-unit cost:* $p_i \geq b_i$ *or* $u_i \geq 0, \forall i \in \mathcal{N}$.

Proof: If tenant i does not reduce its energy consumption, then $f_i = 0$, the lemma holds. Otherwise, if $f_i > 0$, since we sort tenants by per-unit claimed cost, based on the which we select eligible tenants, it is easy to see $\sum_{i \in \mathcal{N}} b_i f_i \leq \sum_{j \in \{\mathcal{N} \setminus i\}} b_j f_j^{-i}$. Thus, we have

$$\begin{aligned} \sum_{j \in \{\mathcal{N} \setminus i\}} b_j f_j + b_i f_i \leq & \sum_{j \in \{\mathcal{N} \setminus i\}} b_j f_j^{-i} \\ b_i f_i \leq & \sum_{j \in \{\mathcal{N} \setminus i\}} b_j f_j^{-i} - \sum_{j \in \{\mathcal{N} \setminus i\}} b_j f_j \\ b_i \leq & \, p_i \end{aligned} \tag{4}$$

We can conclude that Lemma 1 holds. ■

In the following, we first analyze truthfulness in the case that the total amount of energy reduction requests in all the reduction signals ($\sum_{t=1}^{T} s_t$) is large enough, such that all eligible tenants would be asked to reduce energy consumption during the EDR event. In this case, the total amount of energy reduction by all eligible tenants except the last picked eligible tenant is $x_i = f_i = e_i$, that by the last picked eligible tenant is $x_i = f_i \leq e_i$, and that by an ineligible tenants is $x_i = f_i = 0$. Due to Eq. (2), the utility of tenant i is:

$$\begin{aligned} u_i &= p_i f_i - v_i f_i \\ &= \left(\sum_{j \in \{\mathcal{N} \setminus i\}} b_j f_j^{-i} \right) - \left(\sum_{j \in \{\mathcal{N} \setminus i\}} b_j f_j \right) - v_i f_i \end{aligned} \tag{5}$$

where the first item on the right-hand side, $\sum_{j \in \{\mathcal{N} \setminus i\}} b_j f_j^{-i}$, is independent on i (computed when tenant i does not participate in the auction). We therefore can ignore it, but use

$$u_i \sim -\left(\sum_{j\in\{\mathcal{N}\backslash i\}} b_j f_j\right) - v_i f_i \tag{6}$$

(where $\sim$ means that u_i is only related to the right-hand side), in our following proof of Lemmas 2, 3 and 4, which show truthfulness of tenants in the case of large enough $\sum_{t=1}^{T} s_t$. Then we will extend our discussions to the case that not all eligible tenants would be called on to reduce energy consumption during the EDR event, and prove the general truthfulness.

Lemma 2. *An eligible tenant i which reduces energy consumption to its full capacity, i.e., $x_i = f_i = e_i$, during the EDR event cannot increase its utility by misreporting its claimed per-unit cost.*

Proof: For an eligible tenant i which reduces energy consumption to e_i, if it untruthfully claims its cost, it will lead to three cases: **(1)** Tenant i is still an eligible tenant with full energy reduction, then $f_i' = f_i$. According to our *winning tenant allocation* process, it will not affect any f_j, where j represents each tenant other than i; and thus, in Eq. (6) the first item $(-\sum_{j\in\{\mathcal{N}\backslash i\}} b_j f_j)$ remains the same; then, as both v_i and f_i remains, we have $u_i' = u_i$. **(2)** Tenant i becomes an eligible tenant with partial energy reduction (by claiming a higher cost), then $f_i' \leq f_i$, so $\sum_{j\in\{\mathcal{N}\backslash i\}} b_j f_j' > \sum_{j\in\{\mathcal{N}\backslash i\}} b_j f_j$. Given q, this implies some ineligible tenants would have obtained the opportunity to reduce energy consumptions as *new eligible tenants*. Obviously, each of these *new* eligible tenants' per-unit costs is larger than tenant i's. Thus, we have

$$\begin{aligned} \sum_{j\in\{\mathcal{N}\backslash i\}} b_j f_j' - \sum_{j\in\{\mathcal{N}\backslash i\}} b_j f_j &\geq v_i\left(f_i - f_i'\right) \\ \sum_{j\in\{\mathcal{N}\backslash i\}} b_j f_j' + v_i f_i' &\geq \sum_{j\in\{\mathcal{N}\backslash i\}} b_j f_j + v_i f_i \end{aligned} \tag{7}$$

According to Eq. (6), we have $u_i' \leq u_i$. **(3)** Tenant i becomes an ineligible tenant, then $u_i' = 0$. According to Lemma 1 (individual rationality), $u_i \geq 0$. Thus, we conclude that $u_i' \leq u_i$ when an eligible tenant i with full energy reduction misreports its cost. ■

Lemma 3. *An eligible tenant whose energy reduction is less than its full capacity, i.e., $x_i = f_i < e_i$, during the EDR event cannot increase its utility by misreporting its claimed per-unit cost.*

Proof: In our demand response mechanism, we randomly choose q at the beginning, and the auction includes at most one eligible tenant with partial energy reduction; it represents which only owns part of overall energy reduction capacity as eligible. If tenant i misreports its cost, it will lead to three cases: **(1)** Tenant i remains an eligible tenant with partial energy reduction, then $f_i' = f_i$. It is easy to prove $u_i' = u_i$, similarly to the proof of case 1 in Lemma 2. **(2)** Tenant i becomes an eligible tenant with full energy reduction (by claiming a lower cost), then we have $f_i' > f_i$ and $\sum_{j\in\{\mathcal{N}\backslash i\}} b_j f_j' < \sum_{j\in\{\mathcal{N}\backslash i\}} b_j f_j$. The difference between $\sum_{j\in\{\mathcal{N}\backslash i\}} b_j f_j$ and $\sum_{j\in\{\mathcal{N}\backslash i\}} b_j f_j'$ is the total corresponding costs of tenants' energy reductions which are preempted by tenant i. Given q, the amount of preempted energy reductions equals the amount of the tenant i's energy reductions in addition. Obviously, each preempted tenant's per-unit cost is lower than tenant i's. Thus, we have

$$\begin{aligned} \sum_{j\in\{\mathcal{N}\backslash i\}} b_j f_j - \sum_{j\in\{\mathcal{N}\backslash i\}} b_j f_j' &\leq v_i\left(f_i' - f_i\right) \\ \sum_{j\in\{\mathcal{N}\backslash i\}} b_j f_j + v_i f_i &\leq \sum_{j\in\{\mathcal{N}\backslash i\}} b_j f_j' + v_i f_i' \end{aligned} \tag{8}$$

According to Eq. (6), we have $u_i' \leq u_i$. **(3)** Tenant i becomes an ineligible tenant, then $u_i' = 0$. According to Lemma 1, $u_i \geq 0$. Thus, we conclude that $u_i' \leq u_i$ when an eligible tenant i with partial energy reduction misreports its cost. ■

Lemma 4. *An ineligible tenant cannot increase its utility by misreporting its claimed per-unit cost.*

Proof: An ineligible tenant i does not reduce any energy consumption $f_i = 0$; and its utility is zero $u_i = 0$. If an ineligible tenant untruthfully claims its cost (by claiming a lower cost) to join the eligible tenant set, some eligible tenants must reduce less energy consumption, and even become ineligible tenants. Obviously, the average per-unit cost of these losing tenants' energy reductions is less than or equal to the per-unit true cost of tenant i's (otherwise, they may be ineligible when tenant i claims its true cost). Thus, we have

$$\frac{\sum_{j\in\{\mathcal{N}\backslash i\}} b_j f_j - \sum_{j\in\{\mathcal{N}\backslash i\}} b_j f_j'}{f_i'} \leq v_i \tag{9}$$

Based on Eq. (6), the difference between u_i' and u_i is as follows

$$\begin{aligned} u_i' - u_i = &-\left(\sum_{j\in\{\mathcal{N}\backslash i\}} b_j f_j'\right) - v_i f_i' \\ &-\left(-\left(\sum_{j\in\{\mathcal{N}\backslash i\}} b_j f_j\right) - v_i f_i\right) \end{aligned} \tag{10}$$

as $f_i = 0$ and $u_i = 0$, we have

$$u_i' = -\left(\sum_{j\in\{\mathcal{N}\backslash i\}} b_j f_j'\right) - v_i f_i' + \left(\sum_{j\in\{\mathcal{N}\backslash i\}} b_j f_j\right) \tag{11}$$

Based on Eq. (9), we have $u_i' \leq 0$. Thus, we conclude that $u_i' \leq u_i$ when an ineligible tenant i misreports its cost. ■

We next analyze the case that the total amount of energy reduction requests in all the reduction signals $(\sum_{t=1}^{T} s_t)$ is insufficient, such that not all eligible tenants would be called upon for energy reduction throughout the EDR event.

Lemma 5. *In the case that not all eligible tenants are called on for reducing energy consumption during the EDR event, a tenant cannot manipulate its claimed per-unit cost to increase its utility.*

Proof: In this case, due to the uncertainty of reduction signals to arrive, each eligible tenant wishes to get allocated and rewarded at the earlier stage of the EDR event. In our mechanism, we allocate energy reduction in a random permutation of eligible tenants. As we disrupt the order of allocation among them, the allocation order is independent of the claimed costs. Hence, an eligible tenant cannot get allocated earlier by manipulating its claimed cost, which means that an untruthful claimed cost cannot help improve utility; moreover, when it claims a higher cost, it may become an ineligible tenant and its utility decrease to zero.

In some cases, an eligible tenant cannot fully reduce its energy consumption due to its *one time slot energy capacity*, which represents the maximal amount of energy reduction in one time slot. As the capacity c_i is also independent of the claimed cost of eligible tenants i, each tenant cannot reduce more than c_i by manipulating its claimed cost.

In Lemma 4, even if the reduction signals are sufficient for eligible tenants, an ineligible tenant cannot increase its utility by claiming untruthful cost. Thus, we omit the discussion with respect to ineligible tenants here. ■

Finally, we show the truthfulness by FairDR in general.

Theorem 2. *FairDR is a truthful mechanism.*

Proof: Combining Lemmas 2, 3, 4, and 5, it can be easily seen that each tenant has no motivation to manipulate its claimed per-unit cost to improve its utility. According to Definition 1, FairDR is truthful. ■

3) Social cost saving maximization: Without loss of generality, we assume that $v_1, v_2, \dots, v_n$ are sorted in non-decreasing order in our following proofs.

Lemma 6. *In the ideal case that $c_i \geq e_i$, $\forall i \in \mathcal{N}$, i.e., per-time-slot energy reduction of each tenant is only limited by its full capacity, the expected competitive ratio of FairDR is $\frac{\lceil \log_2 m \rceil}{(\frac{1}{2}+\frac{1}{U+4})}$ in social cost saving, where $m = \sum_{i=1}^{N} e_i$ representing the overall energy reduction capacity of all tenants, and $U = \frac{\max_{i \in \mathcal{N}} e_i}{\min_{j \in \mathcal{N}} e_j}$ denoting the maximal ratio of overall energy reduction capacities of any two tenants.*

Proof: We first consider the optimal energy reduction in the ideal case that achieves the maximal social cost saving. In the ideal case, let OPT_k denote the optimal social cost saving; and $OPT_k = \sum_{i=1}^{K-1} (\delta - v_i) e_i + (\delta - v_K) \beta$, subject to $k = \sum_{i=1}^{K-1} e_i + \beta$ and $0 < \beta \leq e_K$ when k units energy reductions to K winning tenants, and β represents the amount of the last eligible tenant's eligible overall energy reduction capacity. $(\delta - v_K) \beta$ represents the social cost saving by last winning tenant's reduction, and $\sum_{i=1}^{K-1} (\delta - v_i) e_i$ represents the first $K-1$ winning tenants' social cost saving. Note that the last winning tenant could be an eligible tenant with full or partial energy reduction. Recall $R = \sum_{t=1}^{T} s_t$. Let OPT_R denote the optimal cost saving of these tenants with lowest per-unit costs for R units of energy reduction. In the step of *eligible tenant initialization*, we randomly choose q from $\{2^1, 2^2, \dots, 2^i, \dots, m\}$ to divide all tenants into eligible and ineligible sets. The probability of choosing each possible value of q is $(1/\lceil \log_2 m \rceil)$. Corresponding to this q, Q denotes the number of eligible tenants; and in the energy reduction allocation process, we make a random permutation among the Q eligible tenants. In the following, it is sufficient to analyze these two specific cases to derive the expected social cost saving.

- Case 1: $\frac{1}{2}R < q \leq R$. In this case, R is sufficient for eligible tenants, and all eligible tenants are called upon; and thus the cost saving equals OPT_q. As the sequence $v_1, v_2, \dots, v_n$ is non-decreasing, the sequence $(\delta - v_1), (\delta - v_2), \dots, (\delta - v_n)$ is non-increasing. According to the non-increasing condition and $\frac{1}{2}R < q \leq R$, the social cost saving is greater than $\frac{1}{2}OPT_R$.
- Case 2: $R < q \leq 2R$. In this case, we select winning tenants from Q eligible tenants. As the last winning tenant might be partially-reduced, we at least select $\lfloor \frac{R}{e_{max}} \rfloor$ winning tenants with full energy reduction. Due to the random selection of eligible tenants, the expectation of social cost saving is greater than $\left(\lfloor \frac{R}{e_{max}} \rfloor / Q \right) OPT_q$. As $R < q \leq 2R$ and we assume that the R units of energy reduction could satisfy at least one tenant ($Q \geq 1$), it is easy to see that

$$\begin{aligned} \frac{\left\lfloor \frac{R}{e_{max}} \right\rfloor}{Q} &\geq \frac{\left\lfloor \frac{R}{e_{max}} \right\rfloor}{\left\lfloor \frac{R}{e_{max}} \right\rfloor + \left\lfloor \frac{R}{e_{min}} \right\rfloor + 2} \\ &\geq \frac{\left\lfloor \frac{R}{e_{max}} \right\rfloor}{\left\lfloor \frac{R}{e_{max}} \right\rfloor + U \left\lfloor \frac{R}{e_{max}} \right\rfloor + 3} \\ &\geq \frac{1}{U+4} \end{aligned} \tag{12}$$

Thereby the social cost saving is greater than $\frac{1}{U+4} OPT_q$.

We denote the social cost saving of Case 1 and Case 2 by SA_1 and SA_2. In Case 1, we have $SA_1 \geq \frac{1}{2}OPT_R$; in Case 2, because $q > R$ and $SA_2 > \frac{1}{U+4}OPT_q$, we have $OPT_q > OPT_R$ and $SA_2 > \frac{1}{U+4}OPT_q > \frac{1}{U+4}OPT_R$. Hence, the proposed mechanism's social cost saving *in expectation* has an guarantee as follows:

$$\begin{aligned} E[SA] &= (1/\lceil \log_2 m \rceil) \cdot \left(\sum_{k=1}^{\lceil \log_2 m \rceil} SA_k \right) \\ &\geq (1/\lceil \log_2 m \rceil) \cdot (SA_1 + SA_2) \\ &\geq (1/\lceil \log_2 m \rceil) \cdot \left(\frac{1}{2}OPT_R + \frac{1}{U+4}OPT_R \right) \\ &= (1/\lceil \log_2 m \rceil) \cdot \left(\frac{1}{2} + \frac{1}{U+4} \right) OPT_R \end{aligned} \tag{13}$$

■

Next, we analyze the expected competitive ratio of FairDR in the general case, where there could exist $c_i < e_i$, for some tenant(s) i. In the general case, the full power reduction capacity of an eligible tenant may not be reached, even if some units of power reduction request in a time slot may have to be fulfilled by the diesel generator. Hence, the cost saving in the general case is no larger than that in the ideal case.

Lemma 7. *In each time slot, the ratio of the cost saving in the general case to that in the ideal case is greater than α, where $\alpha = \min_{i \in \mathcal{N}} \left\{ \frac{c_i}{e_i} \right\}$ represents the minimal ratio of per-time-slot energy reduction capacity to the overall energy reduction capacity among all tenants.*

Proof: To clarify the analysis process, we use $e_{i,t}$ to denote the *remaining* overall energy reduction capacity of tenant i in time slot t, and $x_{i,t}$ to denote the energy reduction allocated to tenant i in time slot t, such that the cost saving in one time slot is $\sum_{i\in\mathcal{E}} (\delta - b_i)\, x_{i,t}$. In time slot t, in that random sequence, for the first tenant i, only if $c_i < s_t < e_{i,t}$, tenant i will could have a gap of allocation result between two cases (*general case* and *ideal case*); for the entire set of tenants, only if the $\sum_{i\in\mathcal{A}} \min\{c_i, e_{i,t}\} < s_t < \sum_{i\in\mathcal{A}} e_{i,t}$, we could leverage diesel generators to fulfil the energy reduction.

For the first tenant i in $\mathcal{A}$, in each time slot t, when the loss happens, the loss is less than $\left(1 - \frac{c_i}{e_{i,t}}\right) e_{i,t}$. Hence in the entire auction, it is easy to see that the loss of tenant i is less than $\left(1 - \frac{c_i}{e_i}\right) e_i$; then, for all tenants, in the entire auction, it is easy to see that the total loss amount of energy reduction is less than $(1-\alpha)$ times of total amount of energy reduction in the ideal case; correspondingly, the total *social cost saving* in general case is larger than α times of it in ideal case. ■

Theorem 3. *FairDR achieves an expected competitive ratio of* $\frac{\lceil \log_2 m \rceil}{\alpha\left(\frac{1}{2}+\frac{1}{U+4}\right)}$ *in social cost saving, where* $m = \sum_{i=1}^{N} e_i$, $U = \frac{\max_{i\in\mathcal{N}} e_i}{\min_{j\in\mathcal{N}} e_j}$, *and* $\alpha = \min_{i\in\mathcal{N}} \left\{\frac{c_i}{e_i}\right\}$.

Proof: In general, due to c_i, the one-time-slot energy reduction capacity, we might need to leverage generators to fulfil some of energy reductions; and when we leverage them, the total social cost saving will decrease; therefore, it is easy to see that the social cost saving of optimal solution is less than that in the ideal case. As the ratio of social cost saving in the general case to the optimum in the ideal case is $\frac{\alpha}{\lceil \log_2 m \rceil}\left(\frac{1}{2} + \frac{1}{U+4}\right)$, the ratio of social cost saving in the general case to the offline optimum, computed based on full knowledge of reduction signals, is no smaller than $\frac{\alpha}{\lceil \log_2 m \rceil}\left(\frac{1}{2} + \frac{1}{U+4}\right)$. Correspondingly, the expected competitive ratio is no larger than $\frac{\lceil \log_2 m \rceil}{\alpha\left(\frac{1}{2}+\frac{1}{U+4}\right)}$. ■

IV. Performance Evaluation

A. Data Sets and Simulation Setup

We simulate a colocation data center located in Ashburn, VA, which is a major data center market served by PJM (a primary transmission organization in the U.S. [14]). In our default setting, the colocation data center includes 5 participating tenants, housing 600, 650, 700, 750, and 800 homogeneous servers, respectively. Each server has a peak power of 250W when busy and a static power of 150W when idle. Hence, the tenants' peak server power ranges between 150kW and 200kW. The diesel generator cost δ is 150\$/MWh based on typical power generation efficiency and the current oil price as of 2015 [15].

1) Energy Reduction Target: Our simulation is based on the PJM's EDR report [14] on January 7, 2014 (due to the severity of the weather condition on that day). The data shows the energy reduction over the entire PJM service area during EDR, rather than for a data center. Therefore, for our evaluation, we scale it down and show in Fig. 3(a) the hourly energy reduction requests for our considered data center throughout an 8-hour EDR event on January 7, 2014 (5am-12pm).

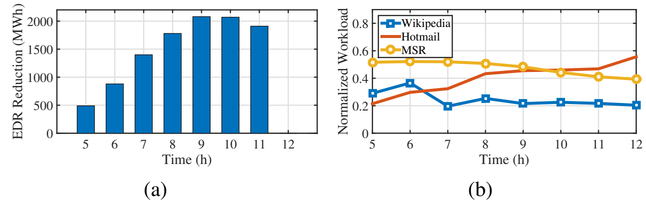

Fig. 3. Trace data. **(a)** Total EDR energy reduction by PJM on January 7, 2014. **(b)** Normalized Workload.

2) Workload: Fig. 3(b) illustrates the workload trace (measuring hourly server utilization) collected from [16], [17]. We assign workload to tenants following these traces. Based on the server power model [16], tenants' total energy consumption is within 1200kWh-1600kWh during the hours spanned by the EDR event, varying according to the workload. In our default setting, the overall energy reduction capacity (e_i) and the maximum energy reduction during each hour (c_i) are 15% of the overall energy consumption during the EDR event at the peak power rate of the respective tenant, and 50% of the per-hour peak power consumption of the respective tenant, while we will vary them later.

3) Tenants' bids: We set the claimed cost of each tenants between $1 \sim 2$ cents/server (equivalently, $0.067 \sim 0.133$\$/kWh), which is comparable to the energy cost saving that could be achieved had the tenants housed servers in their own data centers [16].

Additionally, for computation efficiency of FairDR, we set a minimal energy reduction unit as 10kWh (the typical energy usage of one server rack in one hour), i.e., tenants will migrate the workloads and turn off servers rack-by-rack (subject to performance requirements) when reducing energy consumption. Hence, during each hour, the energy reduction by each tenant is an integral multiple of 10kWh.

B. Results

We compare our mechanism with the *optimal mechanism* and *Truth-DR* [11]. The optimal mechanism, denoted by OPT, achieves the maximal social cost saving by deciding the global optimal allocation with information of all reduction signals. Depending on different rewarding schemes, we introduce two variants of OPT: one is *OPT with Simple Rewards*, which sets the tenants' per-unit rewards equal to their claimed per-unit costs, ignoring the rewarding fairness; another one is *OPT with VCG*, which computes the per-unit rewards by using the VCG pricing scheme [18], [19], [20] in each time slot for ensuring truthfulness in each time slot. Truth-DR is an auction mechanism for colocation EDR developed by Zhang *et al.* [11], which focuses on a one-time auction with a static reduction signal and ensures truthfulness without accounting for fairness. Our evaluation results are shown below.

1) ***Fairness in Rewarding****:* We compare in **Fig. 4** the Coefficients of Variation (CVs) of rewards to winning tenants among OPT, Truth-DR and FairDR, when the EDR event lasts for different numbers of time slots. A data point at a specific T represents the CV computed when the respective mechanism runs for T time slots to allocate energy reduction upon arrivals of the first T reduction signals as shown in Fig. 3(a). Coefficient of variation is a statistical measure of

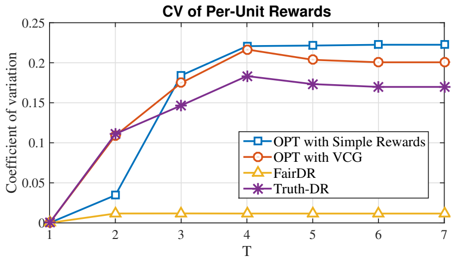

Fig. 4. Comparison of coefficient of variation among different mechanisms.

the dispersion of data points in a data series around the mean, and it is defined as the ratio of the standard deviation σ to the mean μ, *i.e.*, $CV = \frac{\sigma}{\mu}$. As can be seen, despite of the increase of total number of time slots that the EDR event lasts, our mechanism always achieves a CV close to 0, implying that in comparison to OPT and Truth-DR, our mechanism indeed provides the most fair rewarding to tenants.

2) ***Social Cost and Social Cost Saving:*** We further compare the social cost and social cost saving achieved by our mechanism and other mechanisms. For our mechanism with randomization steps, we obtain the expected social cost and social cost saving achieved using all possible values of q. **Fig. 5** and **Fig. 6** show that our mechanism experiences a small increase of social cost and a small loss of social cost saving, as compared to other mechanisms (note that both *OPT with VCG* and *OPT with Simple Rewards* achieve the same cost/cost saving, and hence simply represented by *OPT* in the figures), a tradeoff for our mechanism to provide truthfulness and fairness guarantee in the long-term auction, which the other mechanisms does not provide.

3) ***Competitive Ratio in Social Cost Saving:*** **Figs. 7-10** show the competitive ratio achieved by FairDR in social cost saving by varying different parameters. Especially, in **Fig. 10**, *capacity ratio* represents the ratio of c_i of a tenant to the peak energy consumption of the tenant in one time slot (we set this ratio to be the same for all tenants), where the peak energy consumption is computed assuming all servers of the tenant are running at peak power usage. In **Fig. 10**, *energy ratio* represents the ratio of e_i of a tenant to peak energy consumption of the tenant throughout the EDR (we also set this ratio to be the same for all tenants). We observe that with the increase of the total duration of the EDR event, T, the competitive ratio increases but becomes stable starting from $T = 5$. **Fig. 7** shows that the competitive ratio only increases slightly with the increase of the cost of the diesel generator. **Fig.s 8-10** reveal that the ratio does not change much with different number of participating tenants, and different values of c_i and e_i.

V. Related Work

Data centers' huge yet flexible energy consumption has been increasingly recognized as a valuable demand response resource. In [3], Ghatikar *et al.* verify the feasibility of data center demand response through field tests. In [7], [21], Aikema *et al.* and Ghamkhari *et al.* provide ancillary services, such as voluntary load reduction, for data centers to optimize resource usage, reduce energy cost, even earning additional revenue. Research in [22] optimizes aggregate cost of utilities and data centers via dynamic pricing; and research in [23] also focuses on pricing by investigating data center demand response with prediction-based pricing. All these studies, however, have been focused on owner-operated data centers. A more recent study [24] proposes a *simple* incentive mechanism, called iCODE, for colocation demand response, based on tenants' best-effort reduction for economic demand response without satisfying the mandatory energy reduction as often required by EDR. Further, iCODE cannot ensure truthfulness, and strategic tenants can manipulate their costs to gain extra benefits.

Prior studies [11], [25], [26] have also investigated (data center) demand response based on auction theory. Zhang *et al.* [11] propose a one-time auction mechanism for EDR, Truth-DR, which is computationally efficient, truthful in expectation, and achieves 2-approximation. Zhou *et al.* [25] propose an online procurement auction mechanism in storage-assisted smart grids, which is truthful, computationally efficient, and achieves a constant competitive ratio in practical scenarios. Zhou *et al.* [26] propose an auction mechanism design for demand response in geo-distributed clouds, which is both truthful and computationally efficient. Nonetheless, these studies, except for [11], are not suitable for colocation EDR for which both strategic tenants and operator-controlled diesel generation may commit energy reduction. Moreover, fairness, an important measure in multi-tenant colocations, has been ignored in the prior research.

Previous studies [27], [28], [29] focus on online scenarios where supplies (*e.g.* auction items, perishable goods, idle spectrum channels, power reduction signals) arrive in an online manner, without knowledge of the total supply. Babaioff *et al.* [27] design a mechanism achieving constant approximation when the supply follows a monotone hazard-rate distribution. Goel *et al.* [28] propose an individually-rational, incentive-compatible and Pareto-optimal auction under global budget constraints. Sun *et al.* [29] propose a truthful mechanism achieving channel reusability and fair pricing for spectrum allocation. Our design does not rely on any assumption about the supply.

VI. Concluding Remarks

This paper studies incentive mechanisms for motivating tenants' voluntarily energy consumption reduction in colocation data centers, in the events of emergency demand response. An efficient, truthful auction mechanism, FairDR, is proposed for distributing dynamically arriving energy reduction targets among the tenants and rewarding the tenants for the respective energy cut on the fly. We prove that FairDR provides different tenants similar levels of rewards for the same energy reduction, and guarantees a bounded performance in social cost saving, as compared to the offline optimum. Our simulations based on real-world datasets further validate our analysis and show that FairDR performs consistently well under various settings, compared to other alternative mechanisms. To the best of our knowledge, our work represents the first effort to design online, truthful mechanisms with performance guarantees for colocation demand response, without relying on any *a priori* knowledge of future EDR energy reduction targets. In our future work, we seek to extend our model to other possible scenarios, *e.g.*, one that allows repeated bidding.

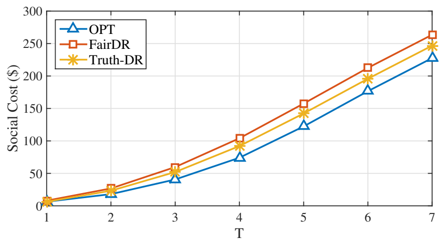

Fig. 5. Comparison of social cost among different mechanisms.

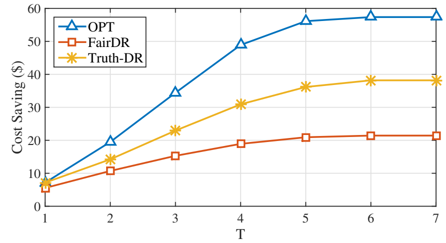

Fig. 6. Comparison of cost saving among different mechanisms.

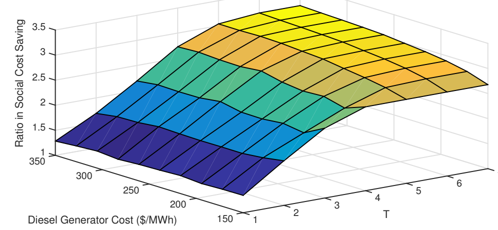

Fig. 7. Competitive ratio in social cost saving: different T, different δ.

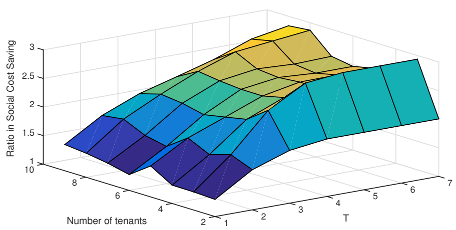

Fig. 8. Competitive ratio in social cost saving: different T, different N.

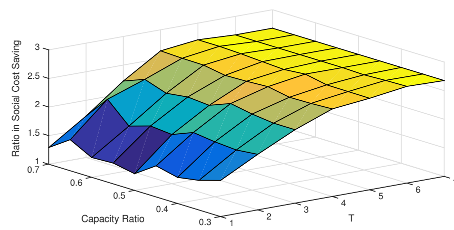

Fig. 9. Competitive ratio in social cost saving: different T, different ratios of c_i to peak per-time-slot power of tenant i.

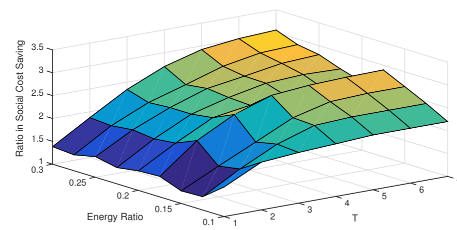

Fig. 10. Competitive ratio in social cost saving: different T, different ratios of e_i to peak overall power of tenant i.

REFERENCES

[1] A. Wierman, Z. Liu, I. Liu, and H. Mohsenian-Rad, “Opportunities and Challenges for Data Center Demand Response,” in *International Green Computing Conference (IGCC)*, Nov 2014, pp. 1–10.

[2] EnerNOC, “Ensuring U.S. Grid Security and Reliability: U.S. EPA’s Proposed Emergency Backup Generator Rule,” 2013.

[3] G. Ghatikar, V. Ganti, N. E. Matson, and M. A. Piette, “Demand Response Opportunities and Enabling Technologies for Data Centers: Findings From Field Studies,” 2012.

[4] Data Center Map, “Colocation USA,” http://www.datacentermap.com/.

[5] L. A. Barroso, J. Clidaras, and U. Hoelzle, *The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines*. Morgan & Claypool, 2013.

[6] Enaxis Consulting, “Pricing Data Center Co-location Services,” 2009, http://enaxisconsulting.com.

[7] D. Aikema, R. Simmonds, and H. Zareipour, “Data Centres in the Ancillary Services Market,” in *International Green Computing Conference (IGCC)*, 2012, pp. 1–10.

[8] K. Managan, “Demand Repsonse: A Market Overview,” 2014, http://enaxisconsulting.com.

[9] PJM, “Emergency Demand Response (Load Management) Performance Report – 2012/2013,” Dec. 2012.

[10] “PJM - Demand Response,” http://www.pjm.com/markets-and-operations/demand-response.aspx.

[11] L. Zhang, S. Ren, C. Wu, and Z. Li, “A Truthful Incentive Mechanism for Emergency Demand Response in Colocation Data Centers,” in *Proc. of IEEE INFOCOM*, 2015.

[12] R. Urgaonkar, B. Urgaonkar, M. J. Neely, and A. Sivasubramaniam, “Optimal Power Cost Management Using Stored Energy in Data Centers,” in *Proc. of ACM SIGMETRICS*, 2011, pp. 221–232.

[13] W. Shi, L. Zhang, C. Wu, Z. Li, and F. C. Lau, “An Online Auction Framework for Dynamic Resource Provisioning in Cloud Computing,” in *Proc. of ACM SIGMETRICS*, 2014, pp. 71–83.

[14] “PJM Interconnection,” http://www.pjm.com/.

[15] Wikipedia, http://en.wikipedia.org/wiki/Diesel_generator.

[16] M. Lin, A. Wierman, L. L. Andrew, and E. Thereska, “Dynamic Right-sizing for Power-proportional Data Centers,” *IEEE/ACM Transactions on Networking (TON)*, vol. 21, no. 5, pp. 1378–1391, 2013.

[17] V. Vasudevan, A. Phanishayee, H. Shah, E. Krevat, D. G. Andersen, G. R. Ganger, G. A. Gibson, and B. Mueller, “Safe and Effective Fine-grained TCP Retransmissions for Datacenter Communication,” in *Proc. of ACM SIGCOMM*, 2009.

[18] W. Vickrey, “Counterspeculation, Auctions, and Competitive Sealed Tenders,” *The Journal of finance*, vol. 16, no. 1, pp. 8–37, 1961.

[19] T. Groves, “Incentives in Teams,” *Econometrica: Journal of the Econometric Society*, pp. 617–631, 1973.

[20] E. Clarke, “Multipart Pricing of Public Goods,” *Public Choice*, vol. 11, no. 1, pp. 17–33, 1971.

[21] M. Ghamkhari and H. Mohsenian-Rad, “Data Centers to Offer Ancillary Services,” in *Proc. of IEEE SmartGridComm*, 2012.

[22] Y. Li, D. Chiu, C. Liu, L. T. Phan, T. Gill, S. Aggarwal, Z. Zhang, B. T. Loo, D. Maier, and B. McManus, “Towards Dynamic Pricing-based Collaborative Optimizations for Green Data Centers,” in *Proc. of IEEE ICDEW*, 2013.

[23] Z. Liu, I. Liu, S. Low, and A. Wierman, “Pricing Data Center Demand Response,” *SIGMETRICS Perform. Eval. Rev.*, vol. 42, no. 1, pp. 111–123, 2014.

[24] S. Ren and M. A. Islam, “Colocation Demand Response: Why do I Turn off My Servers?” in *Proc. of USENIX ICAC*, 2014.

[25] R. Zhou, Z. Li, and C. Wu, “An Online Procurement Auction for Power Demand Response in Storage-Assisted Smart Grids,” in *Proc. of IEEE INFOCOM*, 2015.

[26] 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. of IEEE INFOCOM*, 2015.

[27] M. Babaioff, L. Blumrosen, and A. Roth, “Auctions with Online Supply,” in *Proceedings of the 11th ACM Conference on Electronic Commerce*, ser. EC ’10, 2010.

[28] G. Goel, V. Mirrokni, and R. P. Leme, “Clinching Auctions with Online Supply,” in *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, ser. SODA ’13. SIAM, 2013.

[29] Q. Sun, Q. Wang, K. Ren, and X. Jia, “Fair Pricing in the Sky: Truthful Frequency Allocation with Dynamic Spectrum Supply,” in *Network Protocols (ICNP), 2014 IEEE 22nd International Conference on*, 2014.