

# A Truthful Incentive Mechanism for Emergency Demand Response in Colocation Data Centers

Linquan Zhang\*, Shaolei Ren<sup>†</sup>, Chuan Wu<sup>‡</sup>, Zongpeng Li\*

\*Department of Computer Science, University of Calgary, {linqzhan,zongpeng}@ucalgary.ca

<sup>†</sup> School of Computing and Information Sciences, Florida International University, sren@fiu.edu

<sup>‡</sup>Department of Computer Science, The University of Hong Kong, cwu@cs.hku.hk

**Abstract**—Data centers are key participants in demand response programs, including emergency demand response (EDR), where the grid coordinates large electricity consumers for demand reduction in emergency situations to prevent major economic losses. While existing literature concentrates on owner-operated data centers, this work studies EDR in multi-tenant colocation data centers where servers are owned and managed by individual tenants. EDR in colocation data centers is significantly more challenging, due to lack of incentives to reduce energy consumption by tenants who control their servers and are typically on fixed power contracts with the colocation operator. Consequently, to achieve demand reduction goals set by the EDR program, the operator has to rely on the highly expensive and/or environmentally-unfriendly on-site energy backup/generation. To reduce cost and environmental impact, an efficient incentive mechanism is therefore in need, motivating tenants’ voluntary energy reduction in case of EDR. This work proposes a novel incentive mechanism, Truth-DR, which leverages a reverse auction to provide monetary remuneration to tenants according to their agreed energy reduction. Truth-DR is computationally efficient, truthful, and achieves 2-approximation in colocation-wide social cost. Trace-driven simulations verify the efficacy of the proposed auction mechanism.

## I. INTRODUCTION

For improving the efficiency, reliability and sustainability of power grids, demand response programs are adopted in many countries for exploiting flexibility of electricity usage on the consumer side in response to supply-demand conditions. (see [1] for an overview). Large-scale data centers, given their *large yet flexible* power demands, are widely identified as having a great potential in demand response participation [1], [2]. Notably, data centers have many IT computing knobs (*e.g.*, server speed scaling, workload shedding/migration), as well as non-IT knobs (*e.g.*, battery, cooling system), which are all great assets for demand response to mutually benefit themselves and the power grid [1]. For example, data centers can reduce skyrocketing electricity costs by optimizing workload management, exploiting time-varying and location-dependent electricity prices. Furthermore, data center demand response helps improve power grid efficiency and increase adoption

This work has been supported by Wedge Networks, the Natural Sciences and Engineering Research Council of Canada (NSERC), Hong Kong Research Grant Council (HKU 718513 and HKBU 210412), Alberta Innovates - Technology Futures (AITF) and U.S. National Science Foundation CNS-1423137.

of volatile renewables, by reducing power demand upon low renewable generation, to balance realtime supply-demand [1].

Among various benefits of data center demand response [1], an important, perhaps even the most striking, benefit is the enhancement of power grid reliability through emergency demand response (EDR). In emergency situations (*e.g.*, extreme weather conditions), EDR coordinates many large energy consumers (including data centers) for power demand reduction, and serves as the last line of defense for power grids before cascading blackouts take place, preventing economic losses in the order of *billions of dollars* [3], [4]. In view of the aging grid infrastructure, surging demand and increasingly frequent extreme weather, EDR has undeniably become critically important and seen an upward trend: in PJM, a major regional transmission organization in the U.S., the capacity of EDR power reduction commitment increases from 1,700MW in 2006 – 2007 to 10,800MW in 2011 – 2012 [3]. Because of their huge yet flexible power demand, data centers serve as “energy buffers” and have been identified by U.S. EPA as valuable assets for EDR [2]. As an example, on July 22, 2011, hundreds of data centers participated in EDR and contributed by cutting their electricity usage before a nation-wide blackout occurred in the U.S. and Canada [4].

Existing research on data center demand response heavily concentrates on owner-operated data centers (*e.g.*, Google) [1], [5]. In contrast, we study EDR in a critical yet unique type of data centers, multi-tenant colocation data centers as exemplified by Equinix and often known as “colocation” or “colo”. Unlike owner-operated data centers where operators *fully* manage both servers and facilities, colocation rents physical space out to multiple tenants for housing their own servers, while the colocation operator is mainly responsible for facility support such as cooling and power supply (more details in Sec II). Our study on colocation EDR has a two-fold motivation. *First*, colocations are widely-existing (over 1,200 in the U.S. alone [6]), and according to Google [7], “most large data centers” are colocations. *Second*, many large colocations reside in densely-populated metropolitan areas such as the Silicon Valley [6], where EDR is particularly critical for peak demand reduction, unlike mega-scale owner-operated data centers (*e.g.*, Google) in rural areas with low population densities.

Enabling colocation EDR is challenging because, unlike in owner-operated data centers, colocation power management

is highly “uncoordinated”: the colocation operator purchases electricity from the power grid and manages the facility, whereas individual tenants manage their own servers and power consumption. There is a “split incentive” hurdle: while the colocation operator desires cutting electricity usage for financial compensation received from the power grid in case of EDR, tenants have little incentive for reducing their power consumption, because they are typically billed by the colocation based on their subscribed/reserved power at fixed rates regardless of energy consumption (on top of other power-irrelevant fees such as network connectivity) [8]. To achieve the energy reduction target during EDR, a colocation operator has to resort to highly expensive and/or environmentally-unfriendly energy generation devices (*e.g.*, diesel generators). To reduce the cost and environmental impact of such backup power generation, the tenants should be effectively incentivized to voluntarily cut down their power consumption in case of EDR.

This work proposes a novel incentive mechanism, called Truth-DR, which breaks the split-incentive hurdle by financially rewarding tenants to reduce energy consumption for EDR. Truth-DR is based on a *reverse auction*: upon the notification of an EDR event, tenants *voluntarily* submit bids to specify the amount of their planned energy reduction as well as the associated costs; the colocation operator, as the auctioneer, then decides which bids to accept and the actual monetary reward. Tenants are naturally self-interested and may not truthfully reveal their costs to the colocation operator. Even if tenants are truthful, the problem of deciding the winning bids involves mixed linear integer programming (MLIP) and is NP-hard. Consequently, the Vickrey-Clarke-Groves (VCG) type mechanism [9]–[11], well-known for guaranteeing truthfulness and economical efficiency (social cost minimization), becomes computationally infeasible and is not applicable. Instead, we design Truth-DR based on a randomized auction mechanism, which employs a primal-dual approximation algorithm for winner determination, and strategically assigns rewards to elicit truthful bids. Furthermore, Truth-DR is computationally efficient, individually rational, and guarantees a 2-approximation in colocation-wide social cost, compared with the optimum solution of the NP-hard MLIP problem. We conduct trace-based simulations to validate Truth-DR and corroborate our theoretical analysis, demonstrating the desired efficiency in (social) cost reduction.

The rest of this paper is organized as follows. Related work is reviewed in Section II. The system model and the problem formulation are described in Section III. In Section IV, we present the 2-approximation algorithm and develop our mechanism Truth-DR. Section V provides a simulation study. Finally, concluding remarks are offered in Section VI.

## II. BACKGROUND AND RELATED WORK

- **Colocation** is an integral segment of the data center industry. The now U.S.\$25 billion global colocation market is expected to grow to U.S.\$43 billion by 2018 with a projected annual compound growth rate of 11% [12]. Colocation provides a

unique data center solution to many industry sectors, including leading IT firms such as Twitter, who choose not to construct and maintain private data centers or *completely* outsource IT demands to public cloud providers (due to concerns of privacy and losing control of data). Even top IT companies, such as Microsoft, house some servers in colocations to complement their own data center infrastructure.

Furthermore, colocation is an indispensable enabler for cloud computing, serving as physical homes for many public cloud services offered by small-/medium-scale cloud providers (*e.g.*, Salesforce, Box), which are not “large” enough to construct mega-scale data centers [13]. Last but not least, colocation provides physical support for a significant portion of the Internet traffic, because content delivery network (CDN) providers will handle 55% of Internet traffic by 2018 (up from 36% in 2013) in their servers housed in global colocations in close proximity to user bases [14].

- **Data center demand response.** The critical role of data centers is attracting increasing attention in the field of demand response. Ghatikar *et al.* [5] conduct field tests to verify the feasibility of data center demand response. Ghamkhari *et al.* [15] and Aikema *et al.* [16] optimize data center computing resources to provide ancillary services provided by utility. Studies in [17], [18] investigate the interactions between data centers and utilities, as well as pricing strategies of utilities. The above studies focus on owner-operated data centers where operators have full control over the servers, and are not applicable to colocations. A recent study [13] proposes a *simple* mechanism, iCODE, for colocation demand response, which however cannot be applied to enable *emergency* demand response in colocation data centers, because: (1) iCODE is purely based on tenants’ best-effort reduction that may not meet energy reduction target for EDR; and (2) iCODE is not truthful, and strategic tenants can report falsified costs to gain extra benefits.

- **Auction and its applications.** Zhang *et al.* [19] propose a randomized auction for dynamic virtual machine provisioning in cloud computing. Shi *et al.* [20] further extend it to an online version where the future knowledge is unknown. However, both problems are essentially of a packing type rather than covering type as in this work, and simple extension is not applicable (as shown in Section IV). A decomposition-based randomized auction is proposed for the uncapacitated facility location problem [21]. However, it requires a Lagrangian multiplier preserving approximation algorithm for the underlying problem, which does not exist in our problem. Zhang *et al.* [22] design a reverse auction for the electricity market, nevertheless it is not absolutely truthful.

To our knowledge, this paper represents the *first* step to enable cost-effective EDR in a colocation data center with an efficient and truthful incentive mechanism.

TABLE I: Summary of Notations

Notation	Description	Notation	Description
$N$	# of tenants in colocation	$\delta$	Energy reduction target set by power grid
$\alpha$	BES cost for one unit of power reduction	$\gamma$	PUE (ratio of colocation energy consumption to IT energy consumption)
$n_i$	# of servers turned off by tenant $i$	$f_i$	The payment that the colocation operator provides to a winning tenant $i$
$e_i$	Energy reduction by tenant $i$	$x_i$	Binary variable indicating whether bid $i$ wins or not
$b_i$	Reported cost by tenant $i$	$y$	Continuous variable indicating amount of energy provided by BES

### III. SYSTEM MODEL

#### A. Colocation data center and EDR

Consider a colocation data center with  $N$  tenants, operated by one colocation operator. Each tenant  $i \in \mathcal{N} = \{1, 2, \dots, N\}$  manages its own servers and subscribes a certain amount of power supply from the colocation operator based on a negotiated contract, e.g., 500kW for 24 months at a typical rate of U.S.\$150/kW/month [8]. The colocation operator is responsible for facility management such as cooling. In the event of EDR, a signal is received by the colocation operator from the grid, specifying the amount of energy reduction,  $\delta$ , which the colocation needs to reduce in one EDR period (e.g., one hour [23]) [3]. Such mandatory EDR with a given reduction target is becoming a mainstream approach for EDR to prevent grid-wide blackouts in emergent situations [3].

Colocation power consumption consists of two parts: IT consumption (due to the running servers) controlled by the tenants, and non-IT consumption manageable by the colocation operator. The ratio of the total energy consumption to IT energy consumption is called Power Usage Effectiveness (PUE)  $\gamma$ , which typically ranges from 1.1 to 2.0.

Given power-based contracts, tenants may not actively participate in EDR unless incentivized. Even if some tenants are interested in EDR, their reduction may not collectively reach the energy reduction target  $\delta$ . Hence, the colocation operator often needs to leverage its BES (backup energy storage) to fulfill the shortage of EDR target. Common BES that can produce extra power supply include pre-charged batteries and diesel generators, which are expensive and/or environment unfriendly [16], [24]. Let  $y$  be the amount of grid-power demand reduction due to the usage of BES, and  $\alpha$  be the cost of BES usage per kWh. Such a cost comes from various sources such as wear-and-tear, recharging energy or fuels.

Table I summarizes key notations in the paper.

#### B. Reverse Auction

The proposed solution, Truth-DR, is based on a reverse auction: when EDR signal is issued by the power grid, the colocation operator solicits demand response bids (including the planned energy reduction and associated costs) from tenants.<sup>1</sup> The colocation operator then selects winning bids, decides payments to the winners, and notifies tenants of the auction outcome, as illustrated in Fig. 1.

<sup>1</sup>It is not appealing to directly pass down the grid operator's financial compensation from the colocation operator to tenants, because tenants' responses (i.e., how much energy reduction) are unknown until the end of EDR and consequently, the colocation operator has to primarily rely on its BES for meeting EDR requirement.

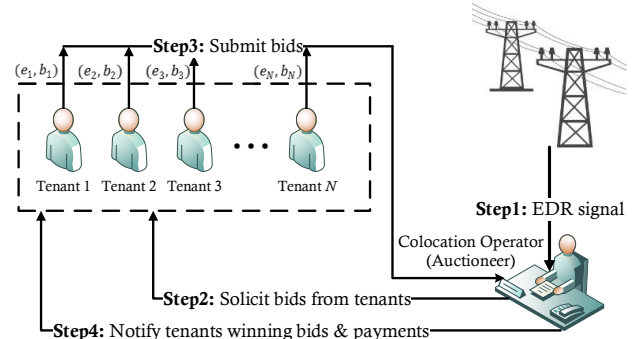


Fig. 1: An illustration of the reverse auction in Truth-DR.

**Bids.** Each tenant  $i$  voluntarily submits a bid to the colocation operator, which specifies two values: (i) planned energy reduction  $e_i$ ; and (ii) claimed cost  $b_i$  due to such a reduction. Each tenant  $i$  has its own discretion to determine its true cost  $c_i$  (which our mechanism will guarantee to be the same as the claimed cost). We will give examples on how each tenant decides its energy reduction and cost in Section V. We also implicitly allow  $(0, 0)$  as a bid, indicating that a tenant is not interested and can be excluded from EDR.

**Winner determination (social cost minimization).** The utility of tenant  $i$  is defined as the payment it receives minus its true cost. The following properties are pursued in our auction design: (i) *Truthfulness (in expectation)*. Bidding true costs is a dominant strategy at the tenants, for maximum utility (in expectation). (ii) *Individual rationality*. A non-negative utility is obtained by each truth-telling tenant participating in the auction. (iii) *Computation efficiency*. The auction should run in polynomial time for winner and payment determination. (iv) *Social cost minimization*. We aim at enabling colocation EDR at the minimum colocation-wide (social) cost, to transform a colocation's formidable power demand into an asset.

Let  $x_i$  be a binary variable indicating whether tenant  $i$ 's bid is successful (1) or not (0), and  $f_i$  be the payment that the colocation operator provides to a winning tenant  $i$ . The social cost in the colocation is the sum of tenants' net costs, i.e., cost due to energy reduction minus award from the colocation operator,  $\sum_{i \in \mathcal{N}} b_i x_i - \sum_{i \in \mathcal{N}} f_i x_i$  (assuming truthful bidding), and the colocation operator's cost in using BES and providing financial awards to the winning tenants, i.e.,  $\alpha y + \sum_{i \in \mathcal{N}} f_i x_i$ . With payments cancelling themselves, the social cost is equivalent to aggregate tenant cost due to energy reduction plus the operator's cost for using BES, i.e.,  $\sum_{i \in \mathcal{N}} b_i x_i + \alpha y$ , which represents the negative impact of energy reduction on colocation operation.

Social cost is a commonly-studied metric in mechanism



design and data center demand response [17]. Minimizing the social cost is equivalent to maximizing the social welfare in our system, considering that the financial compensation paid by the grid operator to the colocation operator for EDR is fixed. Such a financial compensation is typically computed as the compensation rate times the amount of energy reduction. Both the compensation rate and the energy reduction target are typically determined using separate mechanisms [3]. For example, the compensation rate is often determined through a separate market between the power grid and many EDR participants (*e.g.*, data centers), which cannot be manipulated by a single data center. The energy reduction target is also determined well beforehand, *e.g.*, the colocation may commit an energy reduction capacity for EDR three years ahead [3].

We formulate below the social cost minimization problem, referred to as MinCost, which provides the optimal winner determination decisions and usage of BES (*i.e.*, EDR strategies) for the colocation operator, to achieve the energy reduction target  $\delta$ , assuming truthful bids are known.

$$\text{MinCost: minimize}_{x,y} \quad \alpha y + \sum_{i \in \mathcal{N}} b_i x_i \quad (1)$$

subject to:

$$y + \gamma \sum_{i \in \mathcal{N}} e_i x_i \geq \delta, \quad (1a)$$

$$x_i \in \{0, 1\}, \quad \forall i \in \mathcal{N}, \quad (1b)$$

$$y \geq 0. \quad (1c)$$

where  $\gamma \cdot \sum_{i \in \mathcal{N}} e_i x_i$  is the colocation-level energy reduction resulting from tenants' server energy reduction  $\sum_{i \in \mathcal{N}} e_i x_i$ .

MinCost belongs to minimum knapsack problems and is NP-hard [25]. This makes direct application of the VCG mechanism computationally infeasible, for VCG requires exactly solving the underlying winner determination problem multiple times. We instead design a randomized auction that is truthful in expectation and also computationally efficient. Before concluding this section, we note that a baseline, based on average energy consumption of the same time during past weeks, will be used to verify tenants' actual energy reduction, which is typical in incentive-based approaches and similar to how power utility verifies its customer's energy reduction [3].

#### IV. TRUTH-DR: TRUTHFUL INCENTIVE MECHANISM

This section develops a truthful and efficient mechanism, called Truth-DR, to incentivize tenants' participation in EDR. Truth-DR works based on a randomized mechanism that converts a 2-approximation algorithm, which solves the social cost minimization problem, into a truthful and computational efficient auction. We prove that the randomized auction also achieves a 2-approximation ratio in social cost. In what follows, we first design an efficient, 2-approximation algorithm based on a primal-dual technique, and then design the randomized auction using the algorithm as a plug-in module to achieve truthfulness.

##### A. A 2-Approximation Algorithm

To efficiently solve MinCost in (1) for winner determination and BES usage, a natural approach is to relax the integrality constraints (1b) to  $0 \leq x_i \leq 1, \forall i \in \mathcal{N}$ , to obtain a linear program (LP), solve it using standard LP solution techniques, and then round the (possibly) fractional solution  $\mathbf{x}$  to integers. However, the integrality gap, *i.e.*, the ratio between the optimal social cost of (1) to the optimal social cost of the relaxed LP, can be unbounded. For example, consider a case where only one tenant participates in the auction, with  $N = 1, b_1 = 1, e_1 = \alpha > 1, \gamma = 1, \delta = 1$ . The optimal integer solution to (1) is  $(x_1 = 1, y = 0)$ , and the social cost is 1. The relaxed LP however would pick a solution  $(x_1 = 1/\alpha, y = 0)$ , which results in a social cost of  $1/\alpha$ . The integrality gap is hence  $\alpha$ , which is unbounded when  $\alpha \rightarrow \infty$ .

We design an efficient primal-dual algorithm to provide a feasible solution to (1), which provably achieves a 2-approximation in social cost, based on an enhanced LP relaxation of (1). Following the technique of redundant LP constraints [25], [26], we introduce *valid* inequalities that are satisfied by all feasible mixed integer solutions of (1) into the LP relaxation.

Let  $S$  be a subset of bids in  $\mathcal{N}$ . Define  $\Delta(S) = \delta - \gamma \sum_{i \in S} e_i$ , denoting how much energy reduction the colocation still needs to achieve the target  $\delta$ , when all bids in  $S$  are accepted. Let  $e_i(S) = \min\{\gamma e_i, \Delta(S)\}$  denote the contribution of an additional bid  $i$  in making up the discrepancy. The enhanced linear program relaxation (LPR) is:

$$\text{LPR-Primal: minimize } \alpha y + \sum_{i \in \mathcal{N}} b_i x_i \quad (2)$$

subject to:

$$y + \sum_{i \in \mathcal{N} \setminus S} e_i(S) x_i \geq \Delta(S), \quad \forall S \subseteq \mathcal{N} : \Delta(S) > 0 \quad (2a)$$

$$y \geq 0, x_i \geq 0, \quad \forall i \in \mathcal{N} \quad (2b)$$

Constraints in (2a) can be considered as an enumeration of all possible solutions to achieve the energy reduction target. Each constraint in (2a) assumes that all bids in subset  $S, \forall S \subseteq \mathcal{N}$ , are accepted, and limit the solution space to decisions on other bids in  $\mathcal{N} \setminus S$ , to make up for the gap  $\Delta(S) > 0$  to the energy reduction target. Clearly any feasible *mixed integer* solution ( $x_i$  being integer) to LPR (2) is feasible to the original problem (1), and vice versa. We next formulate the dual of LPR (2) by introducing a dual variable  $z(S)$  corresponding to each constraint in (2a).

$$\text{LPR-Dual: maximize } \sum_{S \subseteq \mathcal{N} : \Delta(S) > 0} \Delta(S) z(S) \quad (3)$$

---

**Algorithm 1** A Primal-Dual 2-Approx. Algorithm for (1)

---

```

1: Input:  $(\alpha, \gamma, e, \mathbf{b}, \delta)$ 
2: Output: solution  $(\mathbf{x}, y)$ 
3: //initialization
4:  $(\mathbf{x}, y) = (\mathbf{0}, 0); \mathbf{z} = \mathbf{0}; \mathcal{C} = \emptyset;$ 
5: while  $\Delta(\mathcal{C}) > 0$  do //Update primal/dual variables iteratively
6:   Increase dual variable  $z(\mathcal{C})$  until some dual constraint goes
   tight;
7:   if  $\sum_{S \subseteq \mathcal{N}: \Delta(S) > 0} z(S) = \alpha$  then
8:      $y = \Delta(\mathcal{C});$  break;
9:   end if
10:  if  $\sum_{S \subseteq \mathcal{N}: i \in \mathcal{N} \setminus S, \Delta(S) > 0} e_i(S)z(S) = b_i$  then
11:     $x_i = 1; \mathcal{C} = \mathcal{C} \cup \{i\};$ 
12:  end if
13: end while

```

---

subject to:

$$\sum_{S \subseteq \mathcal{N}: \Delta(S) > 0} z(S) \leq \alpha \quad (3a)$$

$$\sum_{S \subseteq \mathcal{N}: i \in \mathcal{N} \setminus S, \Delta(S) > 0} e_i(S)z(S) \leq b_i \quad \forall i \in \mathcal{N} \quad (3b)$$

$$z(S) \geq 0 \quad \forall S \subseteq \mathcal{N} : \Delta(S) > 0 \quad (3c)$$

Algorithm 1 shows the approximation algorithm, based on the LPR (2) and its dual (3), to derive a feasible, 2-approximate solution to MinCost in (1). The idea of the algorithm is to construct a mixed integer solution to LPR (2) and a feasible solution to its dual (3) iteratively by increasing the dual variable corresponding to the current set of bids to accept,  $\mathcal{C}$ , until the aggregate power from accepted bids in  $\mathcal{C}$  reaches the energy reduction target  $\delta$ .

**Lemma 1.** *Algorithm 1 computes a feasible solution to MinCost in (1) and LPR (2), as well as a feasible solution to dual (3).*

*Proof:* First we examine if the returned solution is feasible to (1). Values in  $\mathbf{x}$  are initialized to 0 and updated to 1 only in the iterations. Thus (1b) is satisfied. Similarly,  $y$  is initialized to 0 and possibly updated to a positive value only, satisfying guaranteeing (1c). For (1a), there are two cases.

*Case 1:* Algorithm 1 stops when  $\Delta(\mathcal{C}) \leq 0$ . According to the definition of  $\Delta(\mathcal{C})$ , we have  $\gamma \sum_{i \in \mathcal{C}} e_i \geq \delta$ , which implies that constraint (1a) is not violated.

*Case 2:* Algorithm 1 exits from line 8. Hence  $y + \gamma \sum_{i \in \mathcal{C}} e_i = \delta$ . Constraint (1a) is respected too.

We can verify that the feasible solution to (1) is feasible to its LPR (2) as well.

We next examine the dual solution  $z(\mathcal{C})$ . Once the algorithm adds tenant  $i$  into the winner set  $\mathcal{C}$ , increasing  $z(\mathcal{C})$  will not change the value of  $\sum_{S \subseteq \mathcal{N}: i \in \mathcal{N} \setminus S, \Delta(S) > 0} e_i(S)z(S)$ , and hence the corresponding constraint in (3b) is respected. Once constraint (3a) goes tight, the algorithm exists and will not

increase any dual variable again. Therefore, the dual solution is feasible to the dual (3) always.  $\square$

**Theorem 1.** *Algorithm 1 is a 2-approximation algorithm to MinCost in (1); it achieves a social cost that is at most 2 times the optimal social cost of (1).*

*Proof:* Let  $OPT$  denote the optimal social cost computed by solving (1) exactly, and  $OPT_{LPR}$  be the optimal social cost computed by its LPR (2). We analyze the algorithm in the following two cases.

*Case 1:* the while loop stops due to  $\Delta(\mathcal{C}) \leq 0$ . The constraint (3a) never goes tight in this case, and hence  $y = 0$ .

$$\begin{aligned} \sum_{i \in \mathcal{N}} b_i x_i &= \sum_{i \in \mathcal{C}} b_i = \sum_{i \in \mathcal{C}} \sum_{S \subseteq \mathcal{N}: i \in \mathcal{N} \setminus S, \Delta(S) > 0} e_i(S)z(S) \\ &= \sum_{S \subseteq \mathcal{N}: \Delta(S) > 0} \sum_{i \in \mathcal{C} \setminus S} e_i(S)z(S) \end{aligned}$$

Note that

$$\begin{aligned} \sum_{i \in \mathcal{C} \setminus S} e_i(S) &\leq \gamma \sum_{i \in \mathcal{C} \setminus \{\omega\}} e_i - \gamma \sum_{i \in S} e_i + e_\omega(S) \\ &< \delta - \gamma \sum_{i \in S} e_i + e_\omega(S) \\ &= \Delta(S) + e_\omega(S) \leq 2\Delta(S) \end{aligned}$$

where  $\omega$  denotes the last tenant added to the solution set  $\mathcal{C}$ .  $\Delta(\mathcal{C} \setminus \omega) > 0$  implies that  $\delta > \gamma \sum_{i \in \mathcal{C} \setminus \{\omega\}} e_i$ . As a result, we have

$$\sum_{i \in \mathcal{N}} b_i x_i \leq \sum_{S \subseteq \mathcal{N}: \Delta(S) > 0} 2z(S)\Delta(S) \leq 2OPT_{LPR} \leq 2OPT$$

*Case 2:* the while loop stops due to the break in line 8. The objective of the LPR then has two parts:  $\alpha y$  and  $\sum_{i \in \mathcal{N}} b_i x_i$ .

$$\alpha y = \Delta(\mathcal{C}) \sum_{S \subseteq \mathcal{N}: \Delta(S) > 0} z(S)$$

Since  $S \subseteq \mathcal{C}, \forall S : \Delta(S) > 0$  and  $z(S) \neq 0$ , we have  $\Delta(S) \geq \Delta(\mathcal{C})$ . Therefore

$$\alpha y \leq \sum_{S \subseteq \mathcal{N}: \Delta(S) > 0} z(S)\Delta(S) \leq OPT_{LPR}.$$

Similar to the analysis in Case 1, we have:

$$\begin{aligned} \sum_{i \in \mathcal{N}} b_i x_i &= \sum_{S \subseteq \mathcal{N}: \Delta(S) > 0} \sum_{i \in \mathcal{C} \setminus S} e_i(S)z(S) \\ &\leq \sum_{S \subseteq \mathcal{N}: \Delta(S) > 0} z(S) \left( \gamma \sum_{i \in \mathcal{C}} e_i - \gamma \sum_{i \in S} e_i \right) \end{aligned}$$

Note that the algorithm exists with  $\Delta(\mathcal{C}) > 0$ , i.e.,  $\delta > \gamma \sum_{i \in \mathcal{C}} e_i$ . We have:

$$\begin{aligned} \sum_{i \in \mathcal{N}} b_i x_i &\leq \sum_{S \subseteq \mathcal{N}: \Delta(S) > 0} z(S) \left( \delta - \gamma \sum_{i \in S} e_i \right) \\ &= \sum_{S \subseteq \mathcal{N}: \Delta(S) > 0} z(S)\Delta(S) = OPT_{LPR} \end{aligned}$$

---

**Algorithm 2** Truth-DR: Truthful Randomized Auction
 

---

## 1: Optimal Fractional Solution

- Solve LPR (2), obtaining optimal BES usage  $y^*$  and optimal fractional winner decisions  $\mathbf{x}^*$ .

## 2: Decomposition into Mixed Integer Solutions

- Decompose the fractional decisions  $(\min\{\beta\mathbf{x}^*, \mathbf{1}\}, \beta y^*)$  to a convex combination of feasible mixed integer solutions  $(\mathbf{x}^l, y^l), l \in \mathcal{I}$ , of (1) using a convex decomposition technique, using Alg. 1 as the separation oracle in the ellipsoid method to solve the primal/dual decomposition LPs.

## 3: Winner Determination and Payment

- Select a mixed integer solution  $(\mathbf{x}^l, y^l)$  from set  $\mathcal{I}$  randomly, using weights of the solutions in the decomposition as probabilities
- Calculate the payment of tenant  $i$  as

$$f_i = \begin{cases} 0 & \text{if } x_i = 0 \\ b_i + \frac{\int_{b_i}^{\alpha\gamma e_i} \min\{2x_i^*(b, b_{-i}), 1\} db}{\min\{2x_i^*(b_i, b_{-i}), 1\}} & \text{otherwise} \end{cases}$$

(min is component-wise minimum in all the above.)

---

Therefore,  $\alpha y + \sum_{i \in \mathcal{N}} b_i x_i \leq 2OPT_{LPR} \leq 2OPT$ .  $\square$

### B. The Randomized Auction

The idea of Truth-DR is as follows. We compute the optimal fractional solution to LPR in (2), and then employ an LP duality based decomposition technique to decompose the fractional solution into a convex combination of feasible, mixed integer solutions to MinCost in (1). The decomposition exploits the covering structure of the MinCost problem, according to a sequence of recent work that originated from theoretical computer science [21], [27], [28]. We then randomly pick one of the mixed integer solutions as the outcome of the auction, using their weights in the convex combination as probabilities. The payments to the winners are computed according to the rule in Theorem 2, which satisfy the sufficient and necessary condition for truthfulness. The auction mechanism is given in Algorithm 2, with details below.

1) *Optimal Fractional Solution*: The optimal fractional solution  $(\mathbf{x}^*, y^*)$  can be computed by solving LPR (2), applying an efficient LP solution technique such as the primal-dual interior-point method.

2) *Convex Decomposition*: The goal of the decomposition is to find  $\nu_l \in [0, 1]$  and a set of mixed integer solutions  $(\mathbf{x}^l, y^l), \forall l \in \mathcal{I}$ , to the MinCost problem in (1), such that  $\sum_{l \in \mathcal{I}} \nu_l \mathbf{x}^l = \mathbf{x}^*$ ,  $\sum_{l \in \mathcal{I}} \nu_l y^l = y^*$ , and  $\sum_{l \in \mathcal{I}} \nu_l = 1$ . In this way, when the randomized auction chooses the  $l^{\text{th}}$  mixed integer solution with probability  $\nu_l$ , a good approximation ratio in social cost in expectation may be achieved, as that achieved by the optimal fractional solution. However, there in fact does not exist a convex combination of the mixed integer solutions, that exactly equals the fractional solution; since otherwise, the expected social cost achieved by these mixed integer solutions equals that achieved by the fractional

solution, contradicting the fact that the fractional solution achieves a lower social cost than any possible mixed integer solution. Therefore, to enable a feasible decomposition, we need to scale up the optimal fractional solution by a certain factor. If there exists an approximation algorithm that solves the underlying winner determination problem with an approximation ratio of  $\beta$ , then we can use  $\beta$  as the scaling factor [28]. In addition,  $\min\{\beta\mathbf{x}^*, \mathbf{1}\}$  should be used to replace  $\beta\mathbf{x}^*$ , to be decomposed into a convex combination of feasible integer solutions  $\mathbf{x}^l$ 's, in order to ensure that the decomposition is feasible. Otherwise, if there exists an entry in vector  $\beta\mathbf{x}^*$  larger than 1, the decomposition is infeasible: since each entry in  $\mathbf{x}^l$  is at most 1, with a convex combination, we have that the left-hand side of (4a)  $\sum_{l \in \mathcal{I}} \nu_l \mathbf{x}^l \leq \mathbf{1}$ . The linear program for this convex decomposition is:

$$\text{maximize } \sum_{l \in \mathcal{I}} \nu_l \quad (4)$$

subject to:

$$\sum_{l \in \mathcal{I}} \nu_l \mathbf{x}^l = \min\{\beta\mathbf{x}^*, \mathbf{1}\} \quad (4a)$$

$$\sum_{l \in \mathcal{I}} \nu_l y^l \leq \beta y^* \quad (4b)$$

$$\sum_{l \in \mathcal{I}} \nu_l \leq 1 \quad (4c)$$

$$\nu_l \geq 0, \forall l \in \mathcal{I} \quad (4d)$$

The BES cost  $\alpha$  is the cost of the colocation operator (auctioneer) rather than tenants (bidders). The exact decomposition in (4a) is to ensure that a tenant's winning probability satisfies the truthfulness condition in Theorem 2. Non-exact decomposition of  $\beta y^*$  will not affect the truthfulness of the auction, and hence (4b) is an inequality rather than an equality.

We can solve the decomposition problem (4) by finding all the possible mixed integer solutions to (1), and then directly solve LP (4) to derive the decomposition weights  $\nu_l$ 's. But there are exponentially many possible mixed integer solutions and hence an exponential number of variables in this LP. We therefore resort to its dual below, where dual variables  $\mu$ ,  $\omega$  and  $\phi$  are associated with primal constraints (4a), (4b) and (4c), respectively:

$$\text{minimize } \beta y^* \omega + \sum_{i=1}^N \min\{\beta x_i^*, 1\} \mu_i + \phi \quad (5)$$

$$y^l \omega + \sum_{i \in \mathcal{N}} x_i^l \mu_i + \phi \geq 1, \quad \forall l \in \mathcal{I} \quad (5a)$$

$$\omega \geq 0, \phi \geq 0 \quad (5b)$$

Though the dual has an exponential number of constraints, the ellipsoid method can be applied to solve it in polynomial-time [27]. The ellipsoid method can obtain an optimal dual solution using a polynomial number of separating hyperplanes.

Algorithm 1 can help find violated dual constraints that act as separating hyperplanes in the ellipsoid method to cut the solution space, and a feasible mixed integer solution to (1) can be derived each time a separating hyperplane is generated. Hence, a polynomial number of candidate solutions to (1) are produced through the process of the ellipsoid method, and the primal decomposition LP (4) can be reduced to a linear program with a polynomial number of variables ( $\nu_l$ 's) corresponding to these solutions. Then we can solve the reduced primal problem in polynomial time. The correctness of the above decomposition method is given in the following lemma:

**Lemma 2.** *The decomposition method correctly obtains a polynomial number of mixed integer solutions  $x^l$ 's to the MinCost problem (1), and convex combination weights  $\nu_l$ 's which achieve the optimal objective value  $\sum_{l \in \mathcal{I}} \nu_l = 1$  for (4), in polynomial time.*

*Proof:* We show  $\sum_{l \in \mathcal{I}} \nu_l = 1$  by proving that the optimal value of (5) is 1. We first observe that the optimal value of the dual is at most 1 since  $(\boldsymbol{\mu} = \mathbf{0}, \omega = 0, \phi = 1)$  is a feasible solution to the dual. For the sake of contradiction, we suppose  $\beta y^* \omega + \sum_{i=1}^N \min\{\beta x_i^*, 1\} \mu_i + \phi < 1$ . The unconstrained variable  $\boldsymbol{\mu}$ , which possibly leads to negative  $\mu_i$ 's, may make the approximation algorithm work inappropriately. Let

$$\mu_i^+ = \begin{cases} \mu_i & \text{if } \mu_i \geq 0 \text{ and } \beta x_i^* \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

By executing the  $\beta$ -approximation algorithm using  $\omega$  as the BES cost and  $\boldsymbol{\mu}^+$  as the bidding prices, we obtain a mixed integer solution  $(\boldsymbol{x}^\pi, y^\pi)$ ,  $\pi \in \mathcal{I}$  that satisfies:

$$\omega y^\pi + \sum_{i=1}^N \mu_i^+ x_i^\pi \leq \beta \omega \tilde{y}^* + \beta \sum_{i=1}^N \mu_i^+ \tilde{x}_i^* \leq \beta \omega y^* + \beta \sum_{i=1}^N \mu_i^+ x_i^* \quad (6)$$

where  $(\tilde{\boldsymbol{x}}^*, \tilde{y}^*)$  is the optimal fractional solution when  $(\omega, \boldsymbol{\mu}^+)$  are used as the BES cost and the bidding prices, respectively.

Due to the covering nature of the underlying problem, we can reduce more energy from tenants to meet the energy reduction target without violating the feasibility. Let

$$\tilde{x}_i^\pi = \begin{cases} x_i^\pi & \text{if } \mu_i \geq 0 \text{ and } \beta x_i^* \leq 1 \\ 1 & \text{otherwise} \end{cases}$$

We therefore have  $\tilde{y}^\pi \leq y^\pi$  since we need less energy from BES to cover the energy reduction request. We also have:

$$\begin{aligned} \sum_{i=1}^N \mu_i \tilde{x}_i^\pi &= \sum_i \mu_i^+ x_i^\pi + \sum_{\substack{i: \mu_i < 0 \\ \text{or } \beta x_i^* > 1}} \mu_i \\ &\leq \sum_i \mu_i^+ x_i^\pi + \sum_{\substack{i: \mu_i < 0 \\ \text{or } \beta x_i^* > 1}} \min\{\beta x_i^*, 1\} \mu_i \end{aligned} \quad (7)$$

Combining Inequalities (6) and (7), we have

$$\begin{aligned} \omega \tilde{y}^\pi + \sum_{i=1}^N \mu_i \tilde{x}_i^\pi &\leq \beta \omega y^* + \beta \sum_{i=1}^N \mu_i^+ x_i^* + \sum_{\substack{i: \mu_i < 0 \\ \text{or } \beta x_i^* > 1}} \min\{\beta x_i^*, 1\} \mu_i \\ &\leq \beta \omega y^* + \sum_{i=1}^N \min\{\beta x_i^*, 1\} \mu_i \leq 1 - \phi \end{aligned}$$

We find a violated constraint in (5a) as  $(\tilde{\boldsymbol{x}}^\pi, \tilde{y}^\pi)$  is a feasible mixed integer solution to (1). A contradiction occurs, thus the optimal value of the dual (5) is 1. Due to the strong duality of linear programs, the optimal value of the decomposition problem (4) is 1 as well, i.e.,  $\sum_{l \in \mathcal{I}} \nu_l = 1$ .

The ellipsoid method [29] is applied to solve the dual problem (5) with an exponential number of constraints in polynomial-time. Algorithm 1 acts as a separation oracle for generating separating hyperplanes for the dual (5). Once an optimal dual solution is obtained, using a polynomial number of hyperplanes, we can convert the primal (4) to an optimization problem with a polynomial number of constraints corresponding to these hyperplanes. As a result, the convex decomposition can be solved in polynomial time.  $\square$

3) *Winner Determination and Payment:* We decide the winners by randomly selecting a mixed integer solution  $(x^l, y^l)$  from the set obtained through the decomposition method, with the weights  $\nu_l$ 's in the decomposition as probabilities. We next design a payment for each winner, such that truthfulness of the auction can be guaranteed.

We have shown that the optimal fractional solution of our covering problem cannot be decomposed into a series of weighted mixed integer solutions through simply scaling it up by a factor, i.e.,  $\beta x^*$ ; instead, we decompose  $\{\beta x^*, \mathbf{1}\}$ . Our new decomposition rules out applications of a fractional-VCG type of payment. However, we discover that our randomized solution to (1) satisfies a set of nice properties that enable us to exploit another route of truthful payment computation.

Let  $P_i(b_i)$  be the probability that tenant  $i$  with bid cost  $b_i$  wins in the auction, and  $b_{-i}$  denote all bids except  $(e_i, b_i)$ . Our auction renders the following results on  $P_i(b_i)$ .

**Lemma 3.** *Given fixed bids  $b_{-i}$  from all other tenants, the probability that tenant  $i$  wins,  $P_i(b_i)$ , is monotonically non-increasing in  $b_i$  with Truth-DR. Moreover,  $\int_0^\infty P_i(b) db < \infty, \forall i \in \mathcal{N}$ .*

*Proof:* Assume  $b'_i \leq b_i$ . Given a fixed  $b_{-i}$ , let  $x_i^*, x_i^{**}$  be the optimal fractional solution for bidder  $i$  when its bid is  $b_i, b'_i$ , respectively. Note that  $x_i \in \{0, 1\}$ , thus  $E[x_i] = P_i(b_i) \times 1 + (1 - P_i(b_i)) \times 0 = P_i(b_i)$ . We then further have  $P_i(b_i) = E[x_i] = \sum_{l \in \mathcal{I}} \nu_l x_i^l = \min\{\beta x_i^*, 1\}$ . We next are going to prove that  $x_i^{**} \geq x_i^*$ .

Let  $f(\boldsymbol{x}, b_i, b_{-i})$  be the value of the objective of the LPR



with bids  $(b_i, b_{-i})$  when the solution is  $\mathbf{x}$ . We have:

$$f(\mathbf{x}^*, b_i, b_{-i}) \leq f(\mathbf{x}^{**}, b_i, b_{-i}) \quad (8)$$

$$f(\mathbf{x}^{**}, b'_i, b_{-i}) \leq f(\mathbf{x}^*, b'_i, b_{-i}) \quad (9)$$

Considering the difference between  $f(\mathbf{x}^*, b_i, b_{-i})$  and  $f(\mathbf{x}^*, b'_i, b_{-i})$ , we have:

$$f(\mathbf{x}^*, b_i, b_{-i}) - f(\mathbf{x}^*, b'_i, b_{-i}) = (b_i - b'_i)x_i^*$$

Similarly,  $f(\mathbf{x}^{**}, b_i, b_{-i}) - f(\mathbf{x}^{**}, b'_i, b_{-i}) = (b_i - b'_i)x_i^{**}$ . For the sake of contradiction, we suppose  $x_i^{**} < x_i^*$ , then

$$f(\mathbf{x}^{**}, b_i, b_{-i}) - f(\mathbf{x}^{**}, b'_i, b_{-i}) < f(\mathbf{x}^*, b_i, b_{-i}) - f(\mathbf{x}^*, b'_i, b_{-i}) \quad (10)$$

Add (9) to the inequality above, we have:

$$f(\mathbf{x}^{**}, b_i, b_{-i}) < f(\mathbf{x}^*, b_i, b_{-i})$$

which contradicts (8). Therefore  $x_i^{**} \geq x_i^*$  and  $P_i(b'_i) \geq P_i(b_i)$ .

In LPR (2), if  $b_i > \alpha\gamma e_i$  then  $x_i^* = 0$ , because if the price offered by the tenant  $i$  is too high, the system will use BES instead. Therefore  $P_i(b_i) = \min\{\beta x_i^*, 1\} = 0$  in this case. We further have

$$\int_0^\infty P_i(b)db = \int_0^{\alpha\gamma e_i} P_i(b)db \leq \alpha\gamma e_i < \infty. \quad \square$$

The sufficient and necessary conditions that we follow to design our payment scheme, to achieve truthfulness, are:

**Theorem 2.** [30] [31] *A randomized auction with bids  $\mathbf{b}$  and payments  $\mathbf{f}$  is truthful in expectation if and only if*

- $P_i(b_i)$  is monotonically non-increasing in  $b_i$ ,  $\forall i \in \mathcal{N}$ ;
- $\int_0^\infty P_i(b)db < \infty$ ,  $\forall i \in \mathcal{N}$ ;
- The expected payment satisfies  $E[f_i] = b_i P_i(b_i) + \int_{b_i}^\infty P_i(b)db$ ,  $\forall i \in \mathcal{N}$ .

We design the payments of winning tenants as

$$f_i = b_i + \frac{\int_{b_i}^{\alpha\gamma e_i} \min\{2x_i^*(b, b_{-i}), 1\}db}{\min\{2x_i^*(b_i, b_{-i}), 1\}} = b_i + \frac{\int_{b_i}^{\alpha\gamma e_i} P_i(b)db}{P_i(b_i)}$$

where  $x_i^*(b_i, b_{-i})$  is the optimal solution of variable  $x_i$  to LPR (2), when tenant  $i$ 's bid cost is  $b_i$ , and others are  $b_{-i}$ .

As we can calculate  $P_i(b_i)$  by solving LPR (2) given  $b_i$ , then the payment  $f_i$  can be calculated numerically.

**Theorem 3.** *Truth-DR in Algorithm 2 runs in polynomial time, is truthful in expectation, individually rational, and achieves 2-approximation in colocation-wide social cost.*

*Proof:* Based on Lemma 3, it is ready to see that each step in Algorithm 2 involves polynomial-time computation only.

According to Lemma 3, Truth-DR satisfies the first two conditions in Theorem 2. Furthermore, we have

$$\begin{aligned} E[f_i] &= (1 - P_i(b_i)) \times 0 + P_i(b_i) \times \left( b_i + \frac{\int_{b_i}^{\alpha\gamma e_i} P_i(b)db}{P_i(b_i)} \right) \\ &= b_i P_i(b_i) + \int_{b_i}^\infty P_i(b)db \end{aligned}$$

The last equality is due to  $\int_{b_i}^{\alpha\gamma e_i} P_i(b)db = \int_{b_i}^\infty P_i(b)db$  as proven in Lemma 3. Hence Truth-DR is truthful in expectation.

The utility of tenant  $i$  when it reports its true cost  $c_i$  is:

Case 1: tenant  $i$  loses, and its utility  $u_i = 0$ ;

Case 2: tenant  $i$  wins, and its utility

$$\begin{aligned} u_i &= f_i - c_i = c_i + \frac{\int_{c_i}^{\alpha\gamma e_i} P_i(b)db}{P_i(c_i)} - c_i \\ &= \frac{\int_{c_i}^{\alpha\gamma e_i} P_i(b)db}{P_i(c_i)} \geq 0 \end{aligned}$$

In both cases, the utility is non-negative, and so Truth-DR is individually rational.

The expected social cost is  $E[\alpha y + \mathbf{b}^T \mathbf{x}]$ , satisfying

$$E[\alpha y + \mathbf{b}^T \mathbf{x}] \leq \beta \alpha y^* + \mathbf{b}^T \min\{\beta \mathbf{x}^*, \mathbf{1}\} \leq \beta OPT_{LPR} \leq \beta OPT.$$

Since we use Algorithm 1 to solve the MinCost problem (1) with an approximation ratio of  $\beta = 2$  according to Theorem 1, Truth-DR achieves a 2-approximation to the optimal social cost in expectation.  $\square$

## V. PERFORMANCE EVALUATION

### A. Data Sets and Simulation Setup

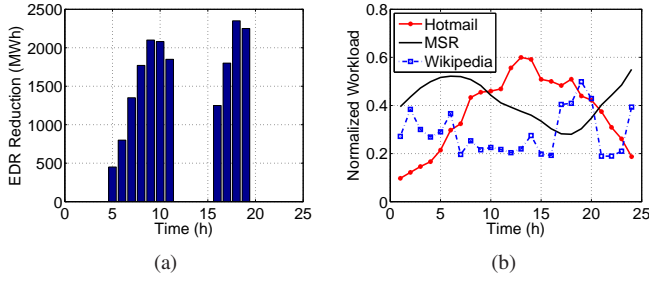
We consider a colocation data center, with six participating tenants (denoted as Tenant #1, Tenant #2,  $\dots$ , and Tenant #6), located at Ashburn, VA, which is a major data center market served by PJM (a major regional transmission organization in the U.S. [23]). Each tenant  $i$  has  $m_i = 10,000$  homogeneous servers with idle/static and computing powers of  $d_{0,i} = 100W$  and  $d_{c,i} = 150W$  each, respectively [13]. The PUE  $\gamma$  of colocation is set to 1.6 (typical for colocation), and the default cost for using BES,  $\alpha$ , is considered \$150/MWh which we will vary later depending on the BES energy source [13], [24]. The peak power demand of the colocation is 24MW.

*Energy reduction targets:* We scale down the total energy reduction by PJM's EDR on January 7, 2014 (when there was severe weather condition) [23], to levels around 15% of the colocation's maximum power, to produce the energy reduction targets in our experiments. We choose 15% because field tests [5] show that data centers can reduce 10-20% energy without affecting normal operation. The total EDR energy reduction by PJM is shown in Fig. 2(a). There were 11 EDR events, starting from 5am to 11am and 16pm to 19pm, respectively, and each event lasted one hour.

*Workload:* We use traces collected from [32] ("Hotmail" and "MSR") and [33] ("Wikipedia"), and, due to limited traces, we duplicate them with randomness of up to 20% to generate the six tenants' workloads. All workloads are normalized with respect to each tenant's maximum service capacity. Fig. 2(b) depicts the three traces.

*Tenants' Energy Reduction bids:* We consider that tenants use the widely-studied knob of "turning off unused servers" to slash energy consumption [32]. In our simulations, the number of servers to turn off by tenant  $i$ ,  $n_i$ , is decided using a widely-considered average queueing delay constraint [13], [32]: Suppose the service rate of a server owned by tenant  $i$  is





**Fig. 2:** Trace data. (a) Total EDR energy reduction by PJM on January 7, 2014. (b) Normalized workloads.

$\mu_i$  (e.g., jobs per unit time) and the workload arrival rate is  $\lambda_i$ ; based on the queueing theory [32], the average job queueing delay is  $\frac{1}{\mu_i - \frac{\lambda_i}{m_i - n_i}}$ , which is required not to exceed a threshold.

We set the parameters according to [13] (e.g., for interactive services,  $\mu = 1,000$ req/second and delay threshold is 20ms). By turning off  $n_i$  servers, total power consumption of tenant  $i$ 's servers becomes  $p'_i = (m_i - n_i) \cdot \left[ d_{0,i} + d_{c,i} \frac{\lambda_i}{(m_i - n_i) \cdot \mu_i} \right]$ , where  $\frac{\lambda_i}{m_i - n_i}$  is the average server utilization (with workloads equally distributed to the servers) [32]. Thus, when no server is turned off (i.e.,  $n_i = 0$ ), the total power consumption is  $p_i = m_i \cdot \left[ d_{0,i} + d_{c,i} \frac{\lambda_i}{m_i \cdot \mu_i} \right]$ , and the total energy reduction by tenant  $i$  is  $e_i = (p_i - p'_i) \cdot T = n_i \cdot d_{0,i} \cdot T$ , where  $T$  is one EDR period, which is 1 hour in our simulations.

*Tenants' Costs:* We consider that tenant  $i$ 's power management cost (e.g., wear-and-tear) for the energy reduction  $e_i$  increases linearly with  $n_i$ , the number of servers turned off, with a slope uniformly distributed between 1 ~ 2 cents/server (equivalently, 6.7 ~ 13.3 cents/kWh). This can reasonably model tenants' costs, because: when tenants house servers in their own data centers, they save 6.7 ~ 13.3 cents/kWh (depending on electricity price), which is naturally enough to cover the power management cost [32].

## B. Results

Our evaluation results are shown below.

1) **Close-to-Minimum Social Cost:** We first compare the social costs achieved by Truth-DR, the optimal integer solution to the MinCost problem in (1), as well as the optimal fractional solution to its LPR in (2), as illustrated in Fig. 3. We obtain the average social cost of our randomized auction by executing Algorithm 2 for 10 times. We observe that Truth-DR provides almost the optimal performance at most time slots, only a slightly higher than the optimum when  $t = 11$ . The results show a close-to-optimum performance in practice, much better than the theoretical performance bound proven in Theorem 1.

2) **Satisfying Energy Reduction Target:** Fig. 4 plots the energy reduction achieved by Truth-DR at each time period, showing that Truth-DR for EDR reaches all energy reduction targets exactly at most time slots except  $t = 11$ , where Truth-DR produces even more reduction than requested. We further

detail the energy reduction provided by each tenant and BES in Fig. 5.

3) **Tenants' Non-Negative Utilities:** Next, we show the payments and utilities received by different tenants in EDR. Fig. 6 illustrates the payments paid by the colocation operator to all tenants. We observe that tenant  $i$  receives no payment when it does not reduce its energy consumption. Fig. 7 shows that all tenants receive non-negative utilities by reducing server energy and participating in EDR, confirming the individual rationality.

4) **Social Cost Reduction Compared to "BES Only":** We show that Truth-DR reduces the colocation-wide social cost compared to using only BES without incentivizing tenants, under different BES costs  $\alpha$  from 150\$/MWh to 350\$/MWh, where 350\$/MWh is the cost for using diesel generators based on typical diesel efficiency [34]. Fig. 8 shows that there is a trend that the social cost saving is more significant as the BES usage gets costlier. Even when  $\alpha = 150$ \$/MWh, there is still a significant saving, showing that Truth-DR enables colocation EDR at a low colocation-wide cost by incentivizing tenants' participation.

Other results, such as tenants' truthfulness (i.e., a tenant's net utility decreases if it claims a falsified cost, as already formally proved in Theorem 3), are omitted for brevity.

## VI. CONCLUSIONS

This work studied how to enable colocation EDR at the minimum colocation-wide cost. To address the challenges of uncoordinated power management and tenants' lack of incentives for EDR, we proposed a first-of-its-kind auction-based incentive mechanism, called Truth-DR, which is computationally efficient, truthful in expectation and guarantees a 2-approximation in colocation-wide social cost. We also performed a trace-driven simulation study to complement the analysis and showed that Truth-DR can achieve the energy reduction target for EDR at a low colocation-wide cost, while ensuring that truthfulness and individual rationality are preserved for tenants during the auction process.

## REFERENCES

- [1] A. Wierman, Z. Liu, I. Liu, and H. Mohsenian-Rad, "Opportunities and Challenges for Data Center Demand Response," in *Proc. of IEEE IGCC*, 2014.
- [2] EnerNOC, "Ensuring U.S. Grid Security and Reliability: U.S. EPA's Proposed Emergency Backup Generator Rule," 2013.
- [3] PJM, "Emergency Demand Response (Load Management) Performance Report - 2012/2013," Dec. 2012.
- [4] A. Misra, "Responding Before Electric Emergencies," <http://www.afcom.com/digital-library/pub-type/communique/responding-before-electric-emergencies/>.
- [5] 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.
- [6] DatacenterMap, "Colocation USA," <http://www.datacentermap.com/>.
- [7] 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.
- [8] Enaxis Consulting, "Pricing Data Center Co-location Services," 2009, <http://enaxisconsulting.com>.
- [9] E. Clarke, "Multipart Pricing of Public Goods," *Public Choice*, vol. 11, no. 1, pp. 17-33, 1971.

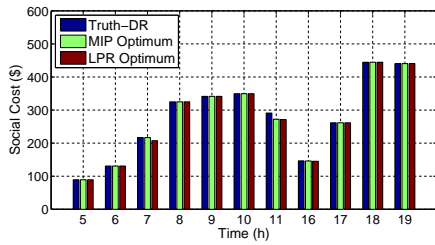


Fig. 3: Comparisons of social cost among different algorithms.

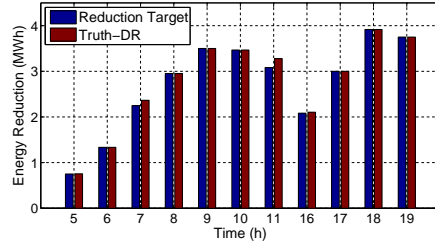


Fig. 4: Comparison of energy reduction target and total energy reduction by Truth-DR

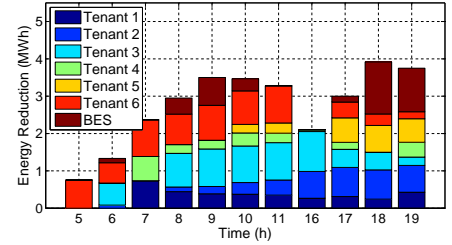


Fig. 5: Comparison of energy reduction by all tenants and BES

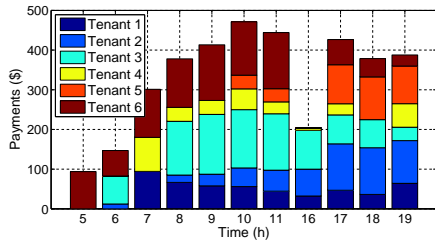


Fig. 6: Comparison of payments received by tenants

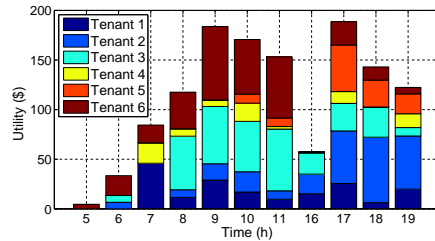


Fig. 7: Comparison of utilities received by tenants

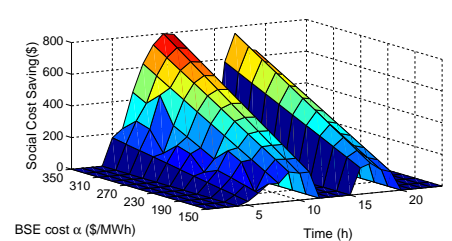


Fig. 8: Social cost saving at different values of BES cost  $\alpha$ .

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

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

[12] "Colocation Market - Worldwide Market Forecast and Analysis (2013 - 2018)," <http://www.marketsandmarkets.com/Market-Reports/colocation-market-1252.html>.

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

[14] Cisco, "Cisco Visual Networking Index: Forecast and Methodology, 2013 – 2018," June 2014.

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

[16] D. Aikema, R. Simmonds, and H. Zareipour, "Data Centres in the Ancillary Services Market," in *Proc. of IEEE IGCC*, 2012.

[17] Z. Liu, I. Liu, S. Low, and A. Wierman, "Pricing Data Center Demand Response," in *Proc. of ACM SIGMETRICS*, 2014.

[18] 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.

[19] L. Zhang, Z. Li, and C. Wu, "Dynamic Resource Provisioning in Cloud Computing: A Randomized Auction Approach," in *Proc. of IEEE INFOCOM*, 2014.

[20] W. Shi, L. Zhang, C. Wu, Z. Li, and F. Lau, "An Online Auction Framework for Dynamic Resource Provisioning in Cloud Computing," in *Proc. of ACM SIGMETRICS*, 2014.

[21] H. Minoei, "Mechanism Design For Covering Problems," 2014.

[22] L. Zhang, Z. Li, and C. Wu, "Randomized Auction Design for Electricity Markets Between Grids and Microgrids," in *Proc. of ACM SIGMETRICS*, 2014.

[23] "PJM Interconnection," <http://www.pjm.com/>.

[24] R. Uргаonkar, B. Uргаonkar, M. J. Neely, and A. Sivasubramaniam, "Optimal Power Cost Management Using Stored Energy in Data Centers," in *Proc. of ACM SIGMETRICS*, 2011.

[25] R. D. Carr, L. K. Fleischer, V. J. Leung, and C. A. Phillips, "Strengthening Integrality Gaps for Capacitated Network Design and Covering Problems," in *Proc. of ACM-SIAM SODA*, 2000.

[26] T. Carnes and D. Shmoys, "Primal-Dual Schema for Capacitated Covering Problems," in *Integer Programming and Combinatorial Optimization*. Springer Berlin Heidelberg, 2008, vol. 5035, pp. 288–302.

[27] R. Carr and S. Vempala, "Randomized Metarounding," *Random Struct. Algorithms*, vol. 20, no. 3, pp. 343–352, May 2002.

[28] R. Lavi and C. Swamy, "Truthful and Near-Optimal Mechanism Design via Linear Programming," in *Proc. of IEEE FOCS*, 2005.

[29] M. Grötschel, L. Lovász, and A. Schrijver, *Geometric Algorithms and Combinatorial Optimization*. Springer-Verlag, 1988.

[30] A. Archer and E. Tardos, "Truthful Mechanisms for One-Parameter Agents," in *Proc. of IEEE FOCS*, 2001.

[31] R. B. Myerson, "Optimal Auction Design," *Mathematics of Operations Research*, vol. 6, no. 1, pp. 58–73, 1981.

[32] M. Lin, A. Wierman, L. L. H. Andrew, and E. Thereska, "Dynamic Right-sizing for Power-proportional Data Centers," in *Proc. of IEEE INFOCOM*, 2011.

[33] 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," *SIGCOMM Comput. Commun. Rev.*, vol. 39, no. 4, pp. 303–314, Aug. 2009.

[34] Wikipedia, [http://en.wikipedia.org/wiki/Diesel\\_generator](http://en.wikipedia.org/wiki/Diesel_generator).