A Truthful Incentive Mechanism for Emergency Demand Response in Geo-distributed Colocation Data Centers

LINQUAN ZHANG, University of Calgary
SHAOLEI REN, University of California, Riverside
CHUAN WU, The University of Hong Kong
ZONGPENG LI, University of Calgary

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 geo-distributed 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.

CCS Concepts: •**Information systems** → **Data centers;** •**Theory of computation** → **Packing and covering problems; Algorithmic mechanism design;** •**Hardware** → **Enterprise level and data centers power issues;**

Additional Key Words and Phrases: Colocation Data Centers, Emergency Demand Response, Mechanism Design, Approximation Algorithms

ACM Reference Format:
Linquan Zhang, Shaolei Ren, Chuan Wu and Zongpeng Li, 2016. A Truthful Incentive Mechanism for Emergency Demand Response in Geo-distributed Colocation Data Centers *ACM Trans. Model. Perform. Eval. Comput. Syst.* 4, 4, Article 1 (June 2016), 22 pages.
DOI: http://dx.doi.org/10.1145/0000000.0000000

1. 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 [Wierman et al. 2014] 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 [EnerNOC 2013; Wierman et al. 2014]. 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, diesel generators), which are all great assets for demand response to mutually benefit themselves and the power grid [Wierman et al. 2014]. For example, data centers can reduce skyrocketing electricity costs by optimizing

This work has been supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), Hong Kong Research Grant Council (HKU 718513, 17204715 and C7036-15G (CRF)), Alberta Innovates - Technology Futures (AITF) and U.S. National Science Foundation (CNS-1551661 and CNS-1565474).
Authors' addresses: L. Zhang and Z. Li, Department of Computer Science, University of Calgary; email: {linqzhan, zongpeng}@ucalgary.ca; S. Ren, Department of Electrical and Computer Engineering, University of California, Riverside; email: sren@ece.ucr.edu; C. Wu, Department of Computer Science, The University of Hong Kong; email: cwu@cs.hku.hk.

DOI: http://dx.doi.org/10.1145/0000000.0000000

workload management, exploiting time-varying and location-dependent electricity prices. Furthermore, data center demand response helps improve power grid efficiency and increase adoption of volatile renewables, by reducing power demand upon low renewable generation, to balance real-time supply-demand.

Among various benefits of data center demand response, 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* [PJM 2012; Misra 2013]. 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 $1,700$MW in $2006-2007$ to $10,800$MW in $2011-2012$ [PJM 2012]. 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 [EnerNOC 2013]. 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 [Misra 2013].

Existing research on data center demand response heavily concentrates on owner-operated data centers (*e.g.*, Google) [Wierman et al. 2014; Ghatikar et al. 2012]. 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 2). Our study on colocation EDR has a two-fold motivation. *First*, colocations are widely-existing (over 1,200 in the U.S. alone [DatacenterMap 2015]), and according to Google [Barroso et al. 2013], "most large data centers" are colocations. *Second*, many large colocations reside in densely-populated metropolitan areas such as the Silicon Valley [DatacenterMap 2015], 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) [Enaxis Consulting gcom]. Further, tenants cannot directly participate in the grid's EDR program by themselves, because the grid can only monitor the colocation's total power (including both IT power and non-IT power) but not individual tenants' power consumption. 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 back-up 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, *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-integer linear programming (MILP) and is NP-hard. Consequently, the Vickrey-Clarke-Groves (VCG) type mechanism [Clarke 1971; Groves 1973; Vickrey 1961], 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 MILP 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 2. The system model and the problem formulation are described in Section 3. In Section 4, we present the 2-approximation algorithm and develop our mechanism Truth-DR. Section 5 provides a simulation study. Finally, concluding remarks are offered in Section 6.

2. 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% [Research and Markets 2013]. 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 [Ren and Islam 2014]. For example, over 40 cloud providers (including VMware) house their servers in a Las Vegas colocation operated by Switch [Data Center Knowledge 2014]. 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 [Cisco 2014].

• **Data center demand response.** The critical role of data centers is attracting increasing attention in the field of demand response. Ghatikar *et al.* [2012] conduct field tests to verify the feasibility of data center demand response. Ghamkhari *et al.* [2012] and Aikema *et al.* [2012] optimize data center computing resources to provide ancillary services provided by utility. Chen *et al.* [2014] leverage server power capping methods to control data center power consumption in response to regulation signals provided by grid operators in real time. Studies in [Liu et al. 2014; Li et al. 2013] 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 [Ren and Islam 2014] 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.

Recently Zhou *et al*. [2015] study the demand response in geo-distributed clouds, and propose an efficient auction, which is truthful with a high probability. Yet their mechanism may be inappropriate for our problem, the EDR in colocations, where 1) such EDR is mandatory with a given energy reduction target; 2) both tenants and colocation's own diesel generators can contribute to such EDR. Sun *et al*. [2015] design an online auction for incentivizing and coordinating tenants' energy reduction in EDR when given tenants' overall energy reduction capacities over time. Never-

Table I: Summary of Notations

Notation	Description
N_k	# of tenants in colo k.
δ_k	Energy reduction target set by power grid for colo k.
α_k	BES cost for one unit of power reduction for colo k.
γ_k	PUE (ratio of colocation energy consumption to IT energy consumption) at colo k.
$\mathcal{K}$	the set of colocations.
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_k	Continuous variable indicating amount of energy provided by BES at colo k.

theless applying their approach to our problem leads to a rather high approximation ratio when total energy reduction is large. Different from their auction, our solution keeps a constant ratio of 2.

• **Auction and its applications.** Auctions are also widely used in computing and communication systems. For instance, Zhang *et al.* [2015] design two online approaches (the pricing approach and the auction approach) for electricity cost saving in colocations when the electricity billing is based on both the total volume consumed and the peak consumption rate. In contrast, our paper focuses on emergency demand response in colocations. Zhang *et al.* [2014a] propose a randomized auction for dynamic virtual machine provisioning in cloud computing. Shi *et al.* [2014] 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 4). A decomposition-based randomized auction is proposed for the uncapacitated facility location problem [Minooei 2014]. However, it requires a Lagrangian multiplier preserving approximation algorithm for the underlying problem, which does not exist in our problem. Zhang *et al.* [2014b] design a reverse auction for the electricity market, nevertheless it is not absolutely truthful. Some auction designs [Baldick et al. 2004] apply supply function bidding to the electricity market, however the supply function bidding suffers from market power issue by large players in the market [Liu et al. 2014].

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.

3. SYSTEM MODEL

3.1. Colocation data centers and EDR

Consider a set of geo-distributed colocation data centers, $\mathcal{K}$, operated by a *colocation operator*. Each colocation $k \in \mathcal{K}$ has N_k *tenants*. Each tenant $i \in \mathcal{N}_k = \{1, 2, \cdots, N_k\}$ 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 [Enaxis Consulting gcom]. 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, δ_k, which the colocation $k \in \mathcal{K}$ needs to cut in one EDR period (*e.g.*, one hour [PJM 2015]) [PJM 2012]. Such mandatory EDR with a given reduction target is becoming a mainstream approach for EDR to prevent grid-wide blackouts in emergent situations [PJM 2012].

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) γ, which typically ranges from 1.1 to 2.0. Let γ_k be the PUE at colocation k.

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 δ. 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 [Urgaonkar et al. 2011; Aikema et al. 2012]. Let y_k be the amount of grid-power demand reduction due to the usage of BES at colocation k, and α_k be the cost of BES usage per kWh at that colocation. Such a cost comes from various sources such as wear-and-tear, recharging energy or fuels.

Table I summarizes key notations in the paper.

3.2. 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. The colocation operator then selects winning bids, decides payments to the winners, and notifies tenants of the auction outcome, as illustrated in Fig. 1. 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. Directly passing down the compensation to tenants also fails to recognize tenants' different performance costs incurred by energy reduction, and therefore may discourage some tenants with high energy reduction related performance costs to participate in EDR.

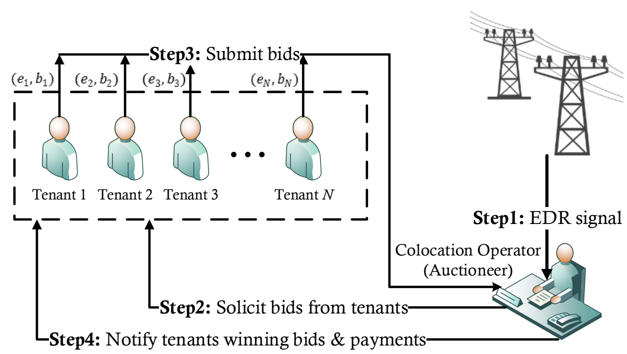

Fig. 1: An illustration of the reverse auction in Truth-DR for one colocation.

Bids. Each tenant $i \in \mathcal{N}_k$ 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 $i \in \mathcal{N}_k$ at colocation k 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 5. 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) *Computational 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 $i \in \mathcal{N}_k$ at colocation k. 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, $\sum_{k\in\mathcal{K},i\in\mathcal{N}_k} b_i x_i - \sum_{k\in\mathcal{K},i\in\mathcal{N}_k} f_i x_i$ (assuming truthful bidding), and the colocation operator's cost in using BES and providing financial awards to the winning tenants, *i.e.*, $\sum_{k\in\mathcal{K}} \alpha_k y_k + \sum_{k\in\mathcal{K},i\in\mathcal{N}_k} f_i x_i$. 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.*, $\sum_{k\in\mathcal{K},i\in\mathcal{N}_k} b_i x_i + \sum_{k\in\mathcal{K}} \alpha_k y_k$, 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 [Liu et al. 2014]. 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 [PJM 2012]. 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 [PJM 2012].

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 δ_k at each colocation $k \in \mathcal{K}$, assuming truthful bids are known.

$$\textbf{MinCost}\text{: minimize}_{\boldsymbol{x},y} \quad \sum_{k\in\mathcal{K}} \alpha_k y_k + \sum_{k\in\mathcal{K},i\in\mathcal{N}_k} b_i x_i \tag{1}$$

subject to:

$$\begin{aligned} y_k + \gamma_k \sum_{i\in\mathcal{N}_k} e_i x_i \geq \delta_k, &\qquad \forall k \in \mathcal{K}, &\qquad (1a) \\ x_i \in \{0,1\}, &\ \ \forall k \in \mathcal{K}, i \in \mathcal{N}_k, &\ (1b) \\ y_k \geq 0. &\qquad \forall k \in \mathcal{K}, &\qquad (1c) \end{aligned}$$

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

MinCost belongs to minimum knapsack problems and is NP-hard [Carr et al. 2000]. 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 [PJM 2012].

4. TRUTH-DR: TRUTHFUL INCENTIVE MECHANSIM

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 computationally efficient auction. We prove that the randomized auction also achieves a 2-approximation ratio in social cost. In what follows, we first decompose MinCost in (1) into a series of sub problem MinCost-ks, and then design an efficient, 2-approximation algorithm based on a primal-dual technique for each sub problem, and next design the randomized auction using the algorithm as a plug-in module to achieve truthfulness.

Let MinCost-k be the sub problem defined as follows for the colocation k:

$$\textbf{MinCost-}k\text{: minimize}_{\boldsymbol{x},y} \quad \alpha_k y_k + \sum_{i \in \mathcal{N}_k} b_i x_i \tag{2}$$

subject to:

$$y_k + \gamma_k \sum_{i \in \mathcal{N}_k} e_i x_i \geq \delta_k, \tag{2a}$$

$$x_i \in \{0,1\}, \ \forall i \in \mathcal{N}_k, \tag{2b}$$

$$y_k \geq 0. \tag{2c}$$

MinCost in (1) then can be decomposed as

$$\textbf{MinCost} = \sum_{k \in \mathcal{K}} \textbf{MinCost-}k \tag{3}$$

4.1. A 2-Approximation Algorithm

For the discussion in this subsection, we shall omit the subscript k for the sake of clarity when there is no confusion based on the context. To efficiently solve MinCost-k in (2) for winner determination and BES usage, a natural approach is to relax the integrality constraints $(2b)$ 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 (2) 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 (2) is $(x_1 = 1, y = 0)$, and the social cost is 1. The relaxed LP however would pick a solution $(x_1 = 1/\alpha_k, y = 0)$, which results in a social cost of $1/\alpha$. The integrality gap is hence α, which is unbounded when $\alpha \to \infty$.

We design an efficient primal-dual algorithm to provide a feasible solution to (2), which provably achieves a 2-approximation in social cost, based on an enhanced LP relaxation of (2). Following the technique of redundant LP constraints [Carr et al. 2000; Carnes and Shmoys 2008], we introduce *valid* inequalities that are satisfied by all feasible mixed integer solutions of (2) 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 δ_k, when all bids in S are accepted. Let $e_i(S) = \min\{\gamma e_i, \Delta(S)\}$ be the contribution of an additional bid i in making up the discrepancy. The enhanced linear program relaxation (LPR) is:

$$\textbf{LPR-Primal:} \quad \text{minimize } \alpha y + \sum_{i \in \mathcal{N}} b_i x_i \tag{4}$$

subject to:

$$y + \sum_{i \in \mathcal{N} \setminus S} e_i(S) x_i \geq \Delta(S), \ \forall S \subseteq \mathcal{N} : \Delta(S) > 0, \tag{4a}$$

$$y \geq 0, x_i \geq 0, \qquad \forall i \in \mathcal{N}. \tag{4b}$$

Constraints in $(4a)$ can be considered as an enumeration of all possible solutions to achieve the energy reduction target. Each constraint in $(4a)$ assumes that all bids in subset $S, \forall S \subseteq \mathcal{N}$, are accepted, and limits 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.

LEMMA 4.1. *Any feasible solution to (2) is feasible to its LPR (4).*

Proof: Let $(y, \boldsymbol{x})$ be a feasible solution to (2). We define $\tilde{S} = \{i | x_i = 1, \forall i \in \mathcal{N}\}$. Then we have following four cases for all constraints $(4a)$.

Case 1: $S \subseteq \tilde{S}$. The left side of $(4a)$ becomes: $y + \sum_{i \in (\mathcal{N} \setminus S) \cap \tilde{S}} \min\{\gamma e_i, \Delta(S)\}$. If $\exists i \in (\mathcal{N} \setminus S) \cap \tilde{S}$, such that $\gamma e_i \geq \Delta(S)$, then:

$$y + \sum_{i \in (\mathcal{N} \setminus S) \cap \tilde{S}} \min\{\gamma e_i, \Delta(S)\} \geq y + \Delta(S) \geq \Delta(S),$$

which implies that constraint $(4a)$ is not violated. If $\forall i \in (\mathcal{N}_k \setminus S) \cap \tilde{S}, \gamma e_i < \Delta(S)$, then:

$$\begin{aligned} y + \sum_{i \in (\mathcal{N} \setminus S) \cap \tilde{S}} \min\{\gamma e_i, \Delta(S)\} &= y + \sum_{i \in (\mathcal{N} \setminus S) \cap \tilde{S}} \gamma e_i \\ &= y + \sum_{i \in \tilde{S}} \gamma e_i - \sum_{i \in \tilde{S} \setminus (\mathcal{N} \setminus S)} \gamma e_i \\ &\geq \delta - \sum_{i \in S} \gamma e_i = \Delta(S), \end{aligned}$$

where the last inequality is due to a) $(y, \boldsymbol{x})$ is a feasible solution to (2); b) $S \subseteq \tilde{S} \Rightarrow \tilde{S} \setminus (\mathcal{N} \setminus S) = S$. So the constraint $(4a)$ is not violated either.

Case 2: $\tilde{S} \subset S$. We have $\mathcal{N} \setminus S \subset \mathcal{N} \setminus \tilde{S}$, which implies that $x_i = 0, \forall i \in \mathcal{N} \setminus S$. The left side of $(4a)$ becomes y. Note that $(y, \boldsymbol{x})$ is a feasible solution to (2), we have

$$\begin{aligned} & y + \gamma \sum_{i \in \tilde{S}} e_i \geq \delta \Rightarrow \\ & y \geq \delta - \gamma \sum_{i \in \tilde{S}} e_i \geq \delta - \gamma \sum_{i \in S} e_i = \Delta(S), \end{aligned}$$

where the last inequality holds due to $\tilde{S} \subset S$.

Case 3: $S \cap \tilde{S} = \emptyset$. The left side of $(4a)$ becomes: $y + \sum_{i \in \tilde{S}} \min\{\gamma e_i, \Delta(S)\}$. If $\exists i \in \tilde{S}$, such that $\gamma e_i \geq \Delta(S)$, then $y + \sum_{i \in \tilde{S}} \min\{\gamma e_i, \Delta(S)\} \geq y + \Delta(S) \geq \Delta(S)$. If $\forall i \in \tilde{S}, \gamma e_i < \Delta(S)$, then $y + \sum_{i \in \tilde{S}} \min\{\gamma e_i, \Delta(S)\} = y + \sum_{i \in \tilde{S}} \gamma e_i \geq \delta \geq \Delta(S)$. Thus the constraint $(4a)$ is respected as well in this case.

Case 4: $S \cap \tilde{S} \neq \emptyset$, $S \nsubseteq \tilde{S}$ and $\tilde{S} \nsubseteq S$. Let $S_1 = S \cap \tilde{S}$, $S_2 = \tilde{S} \setminus S_1$. The left side of $(4a)$ becomes: $y + \sum_{i \in S_2} \min\{\gamma e_i, \Delta(S)\}$. Similarly, the constraint $(4a)$ holds if $\exists i \in S_2$, such that $\gamma e_i \geq \Delta(S)$. If $\forall i \in S_2, \gamma e_i < \Delta(S)$, then:

$$\begin{aligned} y + \sum_{i \in S_2} \min\{\gamma e_i, \Delta(S)\} &= y + \sum_{i \in S_2} \gamma e_i \\ &\geq \delta - \sum_{i \in S_1} \gamma e_i \geq \Delta(S), \end{aligned}$$

Algorithm 1 A Primal-Dual 2-Approximation Algorithm for (2)

```
1: Input: (α, γ, e, b, δ)
2: Output: solution (x, y)
3: //initialization
4: (x, y) = (0, 0); z = 0; C = ∅;
5: while Δ(C) > 0 do //Update primal/dual variables iteratively
6:     Increase dual variable z(C) until some dual constraint goes tight;
7:     if Σ_{S⊆N:Δ(S)>0} z(S) = α then
8:         y = Δ(C); break;
9:     end if
10:    if Σ_{S⊆N:i∈N\S,Δ(S)>0} e_i(S)z(S) = b_i then
11:        x_i = 1; C = C ∪ {i};
12:    end if
13: end while
```

where the last inequality holds due to $S_1 \subsetneq S$. So the constraint $(4a)$ holds as well in this case. □

We next formulate the dual of LPR (4) by introducing a dual variable $z(S)$ corresponding to each constraint in $(4a)$.

$$\textbf{LPR-Dual:} \quad \text{maximize} \sum_{S \subseteq \mathcal{N}:\Delta(S)>0} \Delta(S) z(S) \tag{5}$$

subject to:

$$\sum_{S \subseteq \mathcal{N}:\Delta(S)>0} z(S) \leq \alpha, \tag{5a}$$

$$\sum_{S \subseteq \mathcal{N}: i \in \mathcal{N} \setminus S, \Delta(S)>0} e_i(S) z(S) \leq b_i, \qquad \forall i \in \mathcal{N}, \tag{5b}$$

$$z(S) \geq 0, \ \ \forall S \subseteq \mathcal{N} : \Delta(S) > 0. \tag{5c}$$

Algorithm 1 shows the approximation algorithm, based on the LPR (4) and its dual (5), to derive a feasible, 2-approximate solution to MinCost-k in (2). The idea is to construct a mixed integer solution to LPR (4) and a feasible solution to its dual (5) 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 δ.

LEMMA 4.2. *Algorithm 1 computes a feasible solution to MinCost-k in (2) and LPR (4), as well as a feasible solution to dual (5).*

Proof: First we examine if the returned solution is feasible to (2). Values in $\boldsymbol{x}$ are initialized to 0 and updated to 1 only in the iterations. Thus $(2b)$ is satisfied. Similarly, y is initialized to 0 and possibly updated to a positive value only, satisfying guaranteeing (2c). For (2a), 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 $\gamma_k \sum_{i \in \mathcal{C}} e_i \geq \delta$, which implies that constraint $(2a)$ is not violated.

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

We then verify that the feasible solution to (2) is feasible to its LPR (4) according to Lemma 4.1.

We next examine the dual solution $z(C)$. 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 $(5b)$ is respected. Once constraint $(5a)$ goes tight, the algorithm exists and will not increase any dual variable again. Therefore, the dual solution is feasible to the dual (5) always. □

THEOREM 4.3. *Algorithm 1 is a 2-approximation algorithm to MinCost-k in (2); it achieves a social cost that is at most 2 times the optimal social cost of (2).*

Proof: Let OPT denote the optimal social cost computed by solving (2) exactly, and OPT_{LPR} be the optimal social cost computed by its LPR (4). We analyze the algorithm in the following two cases.

Case 1: the `while` loop stops due to $\Delta(\mathcal{C}) \leq 0$. Let ζ denote the last tenant added to the set $\mathcal{C}$. $z(S) = 0$ if $S \nsubseteq \mathcal{C} \setminus \{\zeta\}$ according to Algorithm 1. The constraint $(5a)$ never goes tight in this case, and hence $y_k = 0$.

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

We are interested in $S \subseteq \mathcal{N}$ whose $z(S) \neq 0$. So we only consider $S \subseteq \mathcal{C} \setminus \{\zeta\}$ in the following analysis.

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

$\Delta(\mathcal{C} \setminus \{\zeta\}) > 0$ implies that $\delta > \gamma \sum_{i \in \mathcal{C} \setminus \{\zeta\}} e_i$, which leads to the second inequality. As a result, we have

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

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

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

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

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

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 exits with $\Delta(\mathcal{C}) > 0$, *i.e.*, $\delta > \gamma \sum_{i \in \mathcal{C}} e_i$. We have:

Algorithm 2 A 2-Approximation Algorithm for (1)

Input: $(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{e}, \boldsymbol{b}, \boldsymbol{\delta})$
Output: solution $(\boldsymbol{x}, \boldsymbol{y})$

for all $k \in \mathcal{K}$ **do**
  if colo k is asked to reduce energy in EDR **then**
    Call Algorithm 1 for colo k with input $(\alpha_k, \gamma_k, \boldsymbol{e}_k, \boldsymbol{b}_k, \delta_k)$;
  end if
end for

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

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

Utilizing Algorithm 1 for each sub problem MinCost-k in (2) as illustrated in Algorithm 2, we then obtain a distributed algorithm for calculating a solution to MinCost in (1). The following theorem unveils that such distributed algorithm produces a 2-approximate solution.

THEOREM 4.4. *Employing Algorithm 1 to solve each sub problem MinCost-k in (2), Algorithm 2 achieves a social cost that is at most 2 times the optimal social cost of MinCost in (1).*

Proof: Let OPT be the optimal social cost computed by solving (1) exactly. Since MinCost can be decomposed as a series of sub problem MinCost-ks, the summation of all LPR-Primal-ks in (4) can act as a linear program relaxation of MinCost in (1). Due to Theorem 4.3, we have

$$\sum_{k \in \mathcal{K}} \alpha_k y_k + \sum_{k \in \mathcal{K}, i \in \mathcal{N}_k} b_i x_i \leq 2 \sum_{k \in \mathcal{K}} OPT_{LPR_k} \leq 2OPT.$$

$\square$

4.2. The Randomized Auction

The idea of Truth-DR is as follows. We compute the optimal fractional solutions to all LPR-Primal-ks in (4), and then employ an LP duality based decomposition technique to decompose the fractional solutions 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 [Carr and Vempala 2002; Lavi and Swamy 2011; Minooei 2014]. 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 4.7, which satisfy the sufficient and necessary condition for truthfulness. The auction mechanism is given in Algorithm 3, with details below.

4.2.1. Optimal Fractional Solution. The optimal fractional solution $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ can be computed by solving all LPR-Primal-ks in (4), applying an efficient LP solution technique such as the primal-dual interior-point method.

4.2.2. Convex Decomposition. The goal of the decomposition is to find $\nu_l \in [0, 1]$ and a set of mixed integer solutions $(\boldsymbol{x}^l, \boldsymbol{y}^l), \forall l \in \mathcal{I}$, to the MinCost problem in (1), such that $\sum_{l \in \mathcal{I}} \nu_l \boldsymbol{x}^l = \boldsymbol{x}^*$, $\sum_{l \in \mathcal{I}} \nu_l \boldsymbol{y}^l = \boldsymbol{y}^*$, and $\sum_{l \in \mathcal{I}} \nu_l = 1$. In this way, when the randomized auction chooses the l^{th} mixed integer solution with probability ν_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

Algorithm 3 Truth-DR: Truthful Randomized Auction

1: Optimal Fractional Solution
— Solve all LPR-Primal-ks in (4), obtaining optimal BES usage $\boldsymbol{y}^*$ and optimal fractional winner decisions $\boldsymbol{x}^*$.
2: Decomposition into Mixed Integer Solutions
— Decompose the fractional decisions $(\min\{\beta\boldsymbol{x}^*, \mathbf{1}\}, \beta\boldsymbol{y}^*)$ to a convex combination of feasible mixed integer solutions $(\boldsymbol{x}^l, \boldsymbol{y}^l), l \in \mathcal{I}$, of (2) using a convex decomposition technique, using Algorithm 2 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 $(\boldsymbol{x}^l, \boldsymbol{y}^l)$ from set $\mathcal{I}$ randomly, using weights of the solutions in the decomposition as probabilities
— Calculate the payment of tenant i as

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

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

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 β, then we can use β as the scaling factor [Lavi and Swamy 2011]. In addition, $\min\{\beta\boldsymbol{x}^*, \mathbf{1}\}$ should be used to replace $\beta\boldsymbol{x}^*$, to be decomposed into a convex combination of feasible integer solutions x^l's, in order to ensure that the decomposition is feasible. Otherwise, if there exists an entry in vector $\beta\boldsymbol{x}^*$ larger than 1, the decomposition is infeasible: since each entry in $\boldsymbol{x}^l$ is at most 1, with a convex combination, we have that the left-hand side of $(6a)$ $\sum_{l\in\mathcal{I}} \nu_l \boldsymbol{x}^l \leq \mathbf{1}$. The linear program for this convex decomposition is:

$$\text{maximize} \quad \sum_{l\in\mathcal{I}} \nu_l \tag{6}$$

subject to:

$$\sum_{l\in\mathcal{I}} \nu_l \boldsymbol{x}^l = \min\{\beta\boldsymbol{x}^*, \mathbf{1}\}, \quad (6a)$$

$$\sum_{l\in\mathcal{I}} \nu_l \boldsymbol{y}^l \leq \beta\boldsymbol{y}^*, \quad (6b)$$

$$\sum_{l\in\mathcal{I}} \nu_l \leq 1, \quad (6c)$$

$$\nu_l \geq 0, \forall l \in \mathcal{I}. \quad (6d)$$

The BES cost $\sum_{k\in\mathcal{K}} \alpha_k y_k$ is the cost of the colocation operator (auctioneer) rather than tenants (bidders). The exact decomposition in $(6a)$ is to ensure that a tenant's winning probability satisfies the truthfulness condition in Theorem 4.7. Non-exact decomposition of $\beta\boldsymbol{y}^*$ will not affect the truthfulness of the auction, and hence $(6b)$ is an inequality rather than an equality.

We can solve the decomposition problem (6) by finding all the possible mixed integer solutions to (2), and then directly solve LP (6) to derive the decomposition weights ν_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 $\boldsymbol{\mu}$, $\boldsymbol{\omega}$ and ϕ are associated with primal constraints $(6a)$, $(6b)$ and $(6c)$, respectively:

$$\text{minimize} \quad \beta \sum_{k\in\mathcal{K}} y_k^* \omega_k + \sum_{k\in\mathcal{K}, i\in\mathcal{N}_k} \min\{\beta x_i^*, 1\}\mu_i + \phi \tag{7}$$

subject to:

$$\begin{aligned} \sum_{k\in\mathcal{K}} y_k^l \omega_k + \sum_{k\in\mathcal{K}, i\in\mathcal{N}_k} x_i^l \mu_i + \phi \geq 1, \qquad & \forall l \in \mathcal{I}, & (7a)\\ \omega_k \geq 0, \phi \geq 0, \qquad & \forall k \in \mathcal{K}, & (7b)\\ \mu_i \text{ unconstrained}, \ & \forall k \in \mathcal{K}, i \in \mathcal{N}_k. & (7c) \end{aligned}$$

Though the dual has an exponential number of constraints, the ellipsoid method can be applied to solve it in polynomial-time [Carr and Vempala 2002]. 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 (2) can be derived each time a separating hyperplane is generated. Hence, a polynomial number of candidate solutions to (2) are produced through the process of the ellipsoid method, and the primal decomposition LP (6) can be reduced to a linear program with a polynomial number of variables (ν_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 4.5. *The decomposition method correctly obtains a polynomial number of mixed integer solutions x^l's to the MinCost problem (2), and convex combination weights ν_l's which achieve the optimal objective value $\sum_{l\in\mathcal{I}} \nu_l = 1$ for (6), in polynomial time.*

Proof: We show $\sum_{l\in\mathcal{I}} \nu_l = 1$ by proving that the optimal value of (7) is 1. We first observe that the optimal value of the dual is at most 1 since $(\boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\omega} = 0, \phi = 1)$ is a feasible solution to the dual. For the sake of contradiction, we suppose the optimal that value of (7) is less than 1, *i.e.*,

$$\beta \sum_{k\in\mathcal{K}} y_k^* \omega_k + \sum_{k\in\mathcal{K}, i\in\mathcal{N}_k} \min\{\beta x_i^*, 1\}\mu_i + \phi < 1.$$

The unconstrained variable $\boldsymbol{\mu}$, which possibly leads to negative μ_is, 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 β-approximation algorithm using $\boldsymbol{\omega}$ as the BES costs and $\boldsymbol{\mu}^+$ as the bidding prices, we obtain a mixed integer solution $(\boldsymbol{x}^\pi, \boldsymbol{y}^\pi), \pi \in \mathcal{I}$ that satisfies:

$$\begin{aligned} \sum_{k\in\mathcal{K}} \omega_k y_k^\pi + \sum_{k\in\mathcal{K}, i\in\mathcal{N}_k} \mu_i^+ x_i^\pi &\leq \beta \sum_{k\in\mathcal{K}} \omega_k \tilde{y}_k^* + \beta \sum_{k\in\mathcal{K}, i\in\mathcal{N}_k} \mu_i^+ \tilde{x}_i^* \\ &\leq \beta \sum_{k\in\mathcal{K}} \omega_k y_k^* + \beta \sum_{k\in\mathcal{K}, i\in\mathcal{N}_k} \mu_i^+ x_i^*, \end{aligned} \tag{8}$$

where $(\tilde{\boldsymbol{x}}^*, \tilde{\boldsymbol{y}}^*)$ is the optimal fractional solution when $(\boldsymbol{\omega}, \boldsymbol{\mu}^+)$ 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:

$$\begin{aligned} \sum_{k\in\mathcal{K}, i\in\mathcal{N}_k} \mu_i \tilde{x_i}^\pi &= \sum_{k\in\mathcal{K}, i\in\mathcal{N}_k} \mu_i^+ x_i^\pi + \sum_{\substack{k\in\mathcal{K},\, i\in\mathcal{N}_k: \\ \mu_i<0 \\ \text{or } \beta x_i^*>1}} \mu_i \\ &\leq \sum_{k\in\mathcal{K}, i\in\mathcal{N}_k} \mu_i^+ x_i^\pi + \sum_{\substack{k\in\mathcal{K},\, i\in\mathcal{N}_k: \\ \mu_i<0 \\ \text{or } \beta x_i^*>1}} \min\{\beta x_i^*, 1\}\mu_i. \end{aligned} \tag{9}$$

Combining Inequalities (8) and (9), we have

$$\begin{aligned} &\sum_{k\in\mathcal{K}} \omega_k \tilde{y}_k^\pi + \sum_{k\in\mathcal{K}, i\in\mathcal{N}_k} \mu_i \tilde{x_i}^\pi \\ &\leq \beta \sum_{k\in\mathcal{K}} \omega_k y_k^* + \beta \sum_{k\in\mathcal{K}, i\in\mathcal{N}_k} \mu_i^+ x_i^* + \sum_{\substack{k\in\mathcal{K},\, i\in\mathcal{N}_k: \\ \mu_i<0 \\ \text{or } \beta x_i^*>1}} \min\{\beta x_i^*, 1\}\mu_i \\ &\leq \beta \sum_{k\in\mathcal{K}} \omega_k y_k^* + \sum_{k\in\mathcal{K}, i\in\mathcal{N}_k} \min\{\beta x_i^*, 1\}\mu_i \leq 1-\phi. \end{aligned}$$

We find a violated constraint in $(7a)$ as $(\tilde{\boldsymbol{x}}^\pi, \tilde{\boldsymbol{y}}^\pi)$ is a feasible mixed integer solution to (2). A contradiction occurs, thus the optimal value of the dual (7) is 1. Due to the strong duality of linear programs, the optimal value of the decomposition problem (6) is 1 as well, *i.e.*, $\sum_{l\in\mathcal{I}} \nu_l = 1$.

The ellipsoid method [Grötschel et al. 1988; Bertsimas and Tsitsiklis 1997] is applied to solve the dual problem (7) with an exponential number of constraints in polynomial-time. In each iteration of the ellipsoid method, Algorithm 1 is called to act as a separation oracle for generating a separating hyperplane for the dual (7). The ellipsoid method can be done in $O\left(n^6 \log(nU)\right)$ iterations [Bertsimas and Tsitsiklis 1997], where n is the number of the decision variables and U is an upper bound on the absolute values of the entries of the constraint matrix. For the dual problem (7), the number of the decision variables is $n' = |\mathcal{K}| + \sum_{k\in\mathcal{K}} |\mathcal{N}_k| + 1$. Both the entries y_k^l and x_k^l are lower bounded by 0 as $(\boldsymbol{x}^l, \boldsymbol{y}^l)$ is a feasible mixed integer solution to the MinCost problem (2). The entries y_k^l is upper bounded by δ_k because the colocation k has to cover at most δ_k energy reduction by using its own BES. The entries x_k^l is upper bounded by 1. So solving the dual problem (7) using ellipsoid method takes $O\left(n'^6 \log(n'U')\right)$ iterations, where $n' = |\mathcal{K}| + \sum_{k\in\mathcal{K}} |\mathcal{N}_k| + 1$ and $U' = \max\{\max_{k\in\mathcal{K}} \delta_k, 1\}$.

Once an optimal dual solution is obtained, using $O\left(n'^6 \log(n'U')\right)$ hyperplanes, we can convert the primal (6) to a linear program with $O\left(n'^6 \log(n'U')\right)$ variables corresponding to these hyperplanes. The current best worst-case complexity algorithm for solving a linear program is given via an interior-point method with a complexity of $O([n^3/\ln n]\gamma)$, where n is the number of decision variables and γ is the number of bits required to represent the linear program [Anstreicher 1999].

For the convex decomposition, there are $O\left(n'^6 \log(n'U')\right)$ variables, leading to a total complexity of $O([m^3/\ln m]\gamma)$, where $m = O\left(n'^6 \log(n'U')\right)$. As a result, the convex decomposition can be solved in polynomial time. □

4.2.3. Winner Determination and Payment. We decide the winners by randomly selecting a mixed integer solution $(\boldsymbol{x}^l, \boldsymbol{y}^l)$ from the set obtained through the decomposition method, with the weights ν_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.*, $\beta\boldsymbol{x}^*$; instead, we decompose $\min\{\beta\boldsymbol{x}^*, \mathbf{1}\}$. Our new decomposition rules out applications of a fractional-VCG type of payment. However, we discover that our randomized solution to (2) 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 4.6. *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 $b'_i \le b_i$. Given a fixed b_{-i}, let x_i^* and x_i^{**} be the optimal fractional solution for bidder i when its bid is b_i and 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 prove that $x_i^{**} \ge x_i^*$.

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 $\boldsymbol{x}$. We have:

$$f(\boldsymbol{x}^*, b_i, b_{-i}) \le f(\boldsymbol{x}^{**}, b_i, b_{-i}) \tag{10}$$

$$f(\boldsymbol{x}^{**}, b'_i, b_{-i}) \le f(\boldsymbol{x}^*, b'_i, b_{-i}) \tag{11}$$

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

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

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

$$f(\boldsymbol{x}^{**}, b_i, b_{-i}) - f(\boldsymbol{x}^{**}, b'_i, b_{-i}) < f(\boldsymbol{x}^*, b_i, b_{-i}) - f(\boldsymbol{x}^*, b'_i, b_{-i}). \tag{12}$$

Add (11) to the inequality above, we have:

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

which contradicts (10). Therefore $x_i^{**} \ge x_i^*$ and $P_i(b'_i) \ge P_i(b_i)$.

In LPR (4), if $b_i > \alpha_k\gamma_k e_i$ for the tenant i in the colocation k 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

$$\int_0^\infty P_i(b)db = \int_0^{\alpha_k\gamma_k e_i} P_i(b)db \le \alpha_k\gamma_k e_i < \infty.$$

□

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

THEOREM 4.7. *[Archer and Tardos 2001][Myerson 1981] A randomized auction with bids* $\boldsymbol{b}$ *and payments* $\boldsymbol{f}$ *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

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

where $x_i^*(b_i, b_{-i})$ is the optimal solution of variable x_i to LPR (4), 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 (4) given b_i, then the payment f_i can be calculated numerically.

THEOREM 4.8. *Truth-DR in Algorithm 3 runs in polynomial time, is truthful in expectation, individually rational, and achieves* 2*-approximation in colocation-wide social cost.*

Proof: Time Complexity. Based on Lemma 4.5, it is ready to see that each step in Algorithm 3 involves polynomial-time computation only.

Truthfulness. According to Lemma 4.6, Truth-DR satisfies the first two conditions in Theorem 4.7. Furthermore, we have

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

The last equality is due to $\int_{b_i}^{\alpha_k \gamma_k e_i} P_i(b)db = \int_{b_i}^{\infty} P_i(b)db$ as proven in Lemma 4.6. Hence, Truth-DR is truthful in expectation.

Individual Rationality. 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

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

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

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

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

Since we use Algorithm 2 to solve the MinCost problem (1) with an approximation ratio of $\beta = 2$ according to Theorem 4.4, Truth-DR achieves a 2-approximation to the optimal social cost in expectation. □

Remarks. The social welfare achieved by the existing randomized auction framework used in cloud computing [Zhang et al. 2014a; Shi et al. 2014] is strictly bounded by the approximation ratio of the plug-in approximation algorithm. Different from their work, the randomized auction in Algorithm 3 could achieve much smaller social cost than the theoretical bound suggested in Theorem 4.8, as observed in the simulation studies in Sec. 5. It is because that the underlying structure of the WDP in our problem is covering type while their problem is packing type. Instead of directly scaling down/up the fractional results in [Zhang et al. 2014a; Shi et al. 2014], we introduce the term $\min\{\beta \boldsymbol{x}^*, \boldsymbol{1}\}$, which has a critical role in successfully decomposing the fractional solution.

5. PERFORMANCE EVALUATION

5.1. Data Sets and Simulation Setup

We consider two colocation data centers, *A* and *B*, each of which has six participating tenants (denoted as Tenant A#1, Tenant A#2, $\cdots$, Tenant A#6, $\cdots$, Tenant B#1, $\cdots$ and Tenant B#6). One colocation *A* is located at Ashburn, VA, which is a major data center market served by PJM (a major regional transmission organization in the U.S. [PJM 2015]). The other colocation *B* is located at New York, NY, which is served by the New York Independent System Operator (NYISO, the operator of New York's electricity bulk grid [NYISO 2015b]). Each tenant i has $m_i = 10,000$ homogeneous servers with idle/static and computing powers of $d_{0,i} = 100$W and $d_{c,i} = 150$W each, respectively [Ren and Islam 2014]. The PUE γ of colocation is set to 1.6 (typical for colocation) for both colocations, and the default cost of both colocations for using BES, α, is considered \$150/MWh which we will vary later depending on the BES energy source [Urgaonkar et al. 2011; Ren and Islam 2014]. The peak power demands of the colocations are 24MW.

Energy reduction targets: For colocation *A*, we scale down the total energy reduction by PJM's EDR on January 7, 2014 (when there was severe weather condition) [PJM 2015], 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 [Ghatikar et al. 2012] 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. There were 11 EDR events, starting from 5am to 11am and 16pm to 19pm, respectively, and each event lasted one hour. For colocation *B*, we use the energy reduction by NYISO's EDR on the same day [NYISO 2015a] as the reduction targets, as illustrated in Fig. 3, which has the same order of magnitude as colocation A's reduction targets.

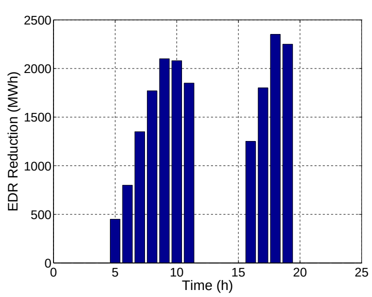

Fig. 2: Total EDR energy reduction by PJM on January 7, 2014.

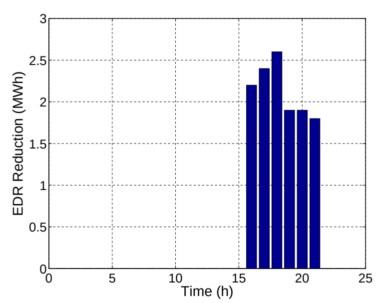

Fig. 3: Total EDR energy reduction by NYISO on January 7, 2014.

Workload: We use traces collected from [Lin et al. 2013] ("Hotmail" and "MSR") and [Vasudevan et al. 2009] ("Wikipedia"), and, due to limited traces, we duplicate them with randomness of up to 20% to generate all the tenants' workloads. All workloads are normalized with respect to each tenant's maximum service capacity. Fig. 4 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 [Lin et al. 2013]. In our simulations, the number of

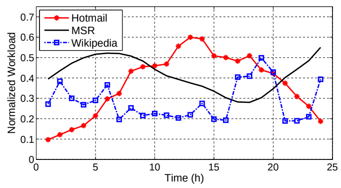

Fig. 4: Normalized workloads from Hotmail, MSR and Wikipedia.

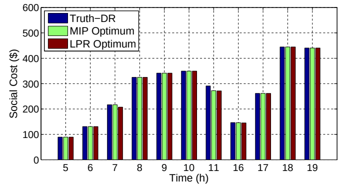

Fig. 5: Comparisons of social cost among different algorithms at colocation *A*.

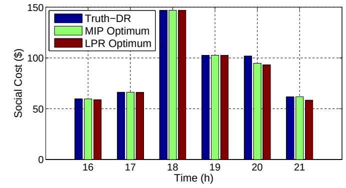

Fig. 6: Comparisons of social cost among different algorithms at colocation *B*.

servers to turn off by tenant i, n_i, is decided using a widely-considered average queueing delay constraint [Lin et al. 2013; Ren and Islam 2014]: Suppose the service rate of a server owned by tenant i is μ_i (*e.g.*, jobs per unit time) and the workload arrival rate is λ_i; based on the queueing theory [Lin et al. 2013], 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 [Ren and Islam 2014] (*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) [Lin et al. 2013]. 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 $1 \sim 2$ cents/server (equivalently, $6.7 \sim 13.3$ cents/kWh). This can reasonably model tenants' costs, because: when tenants house servers in their own data centers, they save $6.7 \sim 13.3$ cents/kWh (depending on electricity price), which is naturally enough to cover the power management cost [Lin et al. 2013].

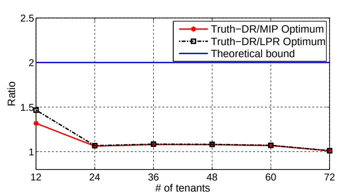

Fig. 7: Comparisons of social cost ratio among different algorithms when the number of tenants varies from 12 to 72.

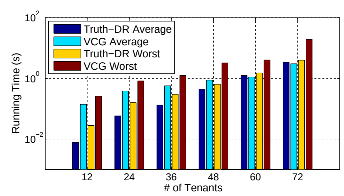

Fig. 8: Comparison of running time among different algorithms when the number of tenants varies.

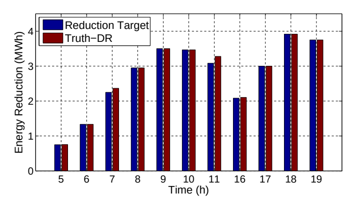

Fig. 9: Comparison of energy reduction target and total energy reduction by Truth-DR at colocation *A*.

5.2. Evaluation Results

5.2.1. **Close-to-Minimum Social Cost**. We first compare the social costs achieved by Truth-DR, the optimal integer solution to the MinCost problem in (2), as well as the optimal fractional solution to its LPR in (4). Results for colocation *A* and *B* are illustrated in Fig. 5 and Fig. 6, respectively. We obtain the average social cost of our randomized auction by executing Algorithm 3 for 10 times. We observe that Truth-DR provides almost-optimal performance in most time slots, only a slightly

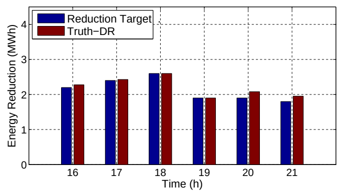

Fig. 10: Comparison of energy reduction target and total energy reduction by Truth-DR at colocation *B*.

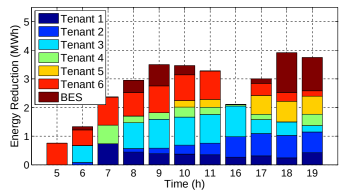

Fig. 11: Comparison of energy reduction by all tenants and BES at colocation *A*.

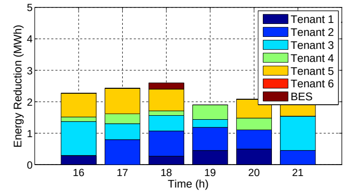

Fig. 12: Comparison of energy reduction by all tenants and BES at colocation *B*.

higher than the optimum when $t = 11$ for colocation *A* and $t = 20$ for colocation *B*. The results show a close-to-optimum performance in practice, much better than the theoretical performance bound proven in Theorem 4.3. We also compare the social costs achieved by Truth-DR, the optimal integer solution and the optimal fractional solution when the number of tenants varies from 12 to 72, as illustrated in Figure 7. A close-to-optimum performance is observed again for all cases. The worst performance, whose approximation ratio is around 1.3, is found when there are 12 tenants. All results suggest that Truth-DR can achieve better performance than the theoretical bound.

5.2.2. **Running Time Comparison**. We are also interested in the running time of Truth-DR. The number of tenants is from 12 to 72. Given a number of tenants, we generate several EDR events, tenants' workloads as well as their energy reduction bids to drive both Truth-DR and VCG. Truth-DR and VCG auction are evaluated using these EDR events. Results are illustrated in Figure 8. It suggests that the average running time of Truth-DR is better than the VCG average in most cases. The running time of both is increasing as the number of tenants increases. We also compare the worst running time by Truth-DR and VCG, and find that the VCG auction, which requires optimally solving the underlying mixed integer program, could perform poorly in some cases, *e.g.*, the running time reaches $19s$ when $N = 72$ while Truth-DR's worst running time is still around $3s$. This result indicates that the NP-complete nature of the underlying problem could make it time consuming to obtain the optimal solution for some cases when the number of tenants is large.

5.2.3. **Satisfying Energy Reduction Target**. Fig. 9 and Fig. 10 plot the energy reduction achieved by Truth-DR at each time period for two colocations, respectively. They show that Truth-DR for EDR reaches all energy reduction targets exactly in most time slots except $t = 11$ for colocation *A* and $t = 20$ as well as $t = 21$ for colocation *B*, where Truth-DR produces even more reduction than requested. We further detail the energy reduction provided by each tenant and BES in Fig. 11 and Fig. 12. We also show the payments and utilities received by different tenants in EDR. Fig. 13 and Fig. 14 illustrate the payments paid by the colocation operator to all tenants in both colocations. We observe that tenant i receives no payment when it does not reduce its energy consumption. Results suggest that more energy reduction implies more payment received by tenants. The relation between energy reduction and payment is not simply linear, but more complicated, reflecting the market supply and demand over time. For instance, the amount of energy reduction by tenant 3 at $T = 17h$ and $T = 18$ at colocation B are almost the same, but the payments are significantly different.

5.2.4. **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 α from 150\$/MWh to 350\$/MWh, where 350\$/MWh is the cost for using diesel generators based on typical diesel efficiency. Fig. 15 shows that there is a trend that the social cost saving is more significant as the BES usage gets costlier. In particular, the cost saving percentage reaches 80% when $T = 16$ and $\alpha = 350$\$/MWh. There is still a significant saving, around 20% to

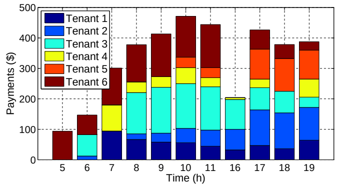

Fig. 13: Comparison of payments received by tenants at colocation *A*.

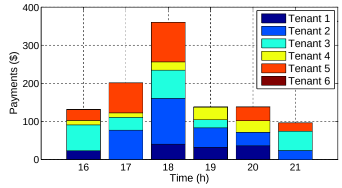

Fig. 14: Comparison of payments received by tenants at colocation *B*.

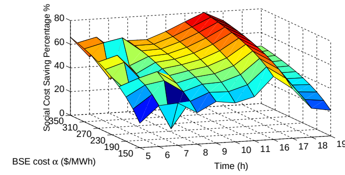

Fig. 15: Social cost saving at different values of BES cost α at colocation *A*.

60% compared with "BES Only", even when $\alpha = 150$\$/MWh. This result indicates that Truth-DR enables colocation EDR at a low colocation-wide cost by incentivizing tenants' participation.

6. 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

David Aikema, Rob Simmonds, and Hamidreza Zareipour. 2012. Data Centres in the Ancillary Services Market. In *Proc. of IEEE IGCC*. 1–10.

K. M. Anstreicher. 1999. Linear Programming in $O([n3/\ln n]L)$ Operations. *SIAM Journal on Optimization* 9, 4 (1999), 803–812.

A Archer and E. Tardos. 2001. Truthful Mechanisms for One-Parameter Agents. In *Proc. of IEEE FOCS*. 482–491.

Ross Baldick, Ryan Grant, and Edward Kahn. 2004. Theory and Application of Linear Supply Function Equilibrium in Electricity Markets. *Journal of Regulatory Economics* 25, 2 (2004), 143–167.

Luiz Andre Barroso, Jimmy Clidaras, and Urs Hoelzle. 2013. *The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines*. Morgan & Claypool.

Dimitris Bertsimas and John N. Tsitsiklis. 1997. *Introduction to Linear Optimization*. Athena Scientific.

Tim Carnes and David Shmoys. 2008. Primal-Dual Schema for Capacitated Covering Problems. In *Integer Programming and Combinatorial Optimization*. Vol. 5035. Springer Berlin Heidelberg, 288–302.

Robert Carr and Santosh Vempala. 2002. Randomized Metarounding. *Random Structures & Algorithms* 20, 3 (May 2002), 343–352.

Robert D. Carr, Lisa K. Fleischer, Vitus J. Leung, and Cynthia A. Phillips. 2000. Strengthening Integrality Gaps for Capacitated Network Design and Covering Problems. In *Proc. of ACM-SIAM SODA*. 106–115.

Hao Chen, Michael C. Caramanis, and Ayse K. Coskun. 2014. The Data Center as a Grid Load Stabilizer. In *Proc. of ASP-DAC*. 105–112.

Cisco. 2014. Cisco Visual Networking Index: Forecast and Methodology, 2013 – 2018. (June 2014).

EdwardH. Clarke. 1971. Multipart Pricing of Public Goods. *Public Choice* 11, 1 (1971), 17–33.

Data Center Knowledge. 2014. Inside SuperNAP 8: Switch's Tier IV Data Fortress. *http://www.datacenterknowledge.com/archives/2014/02/11/inside-supernap-8-switchs-tier-iv-data-fortress/* (2014).

DatacenterMap. 2015. Colocation USA. *http://www.datacentermap.com/* (2015).

Enaxis Consulting. 2009, http://enaxisconsulting.com. Pricing Data Center Co-location Services. (2009, http://enaxisconsulting.com).

EnerNOC. 2013. Ensuring U.S. Grid Security and Reliability: U.S. EPA's Proposed Emergency Backup Generator Rule. (2013).

Mahdi Ghamkhari and Hamed Mohsenian-Rad. 2012. Data Centers to Offer Ancillary Services. In *Proc. of IEEE SmartGridComm*. 436–441.

Girish Ghatikar, Venkata Ganti, Nance Matson, and Mary Ann Piette. 2012. Demand Response Opportunities and Enabling Technologies for Data Centers: Findings From Field Studies. (2012).

Martin Grötschel, László Lovász, and Alexander Schrijver. 1988. *Geometric Algorithms and Combinatorial Optimization*. Springer-Verlag.

Theodore Groves. 1973. Incentives in Teams. *Econometrica: Journal of the Econometric Society* (1973), 617–631.

Ron Lavi and Chaitanya Swamy. 2011. Truthful and Near-Optimal Mechanism Design via Linear Programming. *J. ACM* 58, 6 (2011), 25:1–25:24.

Yang Li, David Chiu, Changbin Liu, Linh T.X. Phan, Tanveer Gill, Sanchit Aggarwal, Zhuoyao Zhang, Boon Thau Loo, David Maier, and Bart McManus. 2013. Towards Dynamic Pricing-based Collaborative Optimizations for Green Data Centers. In *Proc. of IEEE ICDEW*. 272–278.

Minghong Lin, Adam Wierman, Lachlan L. H. Andrew, and Eno Thereska. 2013. Dynamic Right-sizing for Power-proportional Data Centers. *IEEE/ACM Transactions on Networking* 21, 5 (Oct. 2013), 1378–1391.

Zhenhua Liu, Iris Liu, Steven Low, and Adam Wierman. 2014. Pricing Data Center Demand Response. In *Proc. of ACM SIGMETRICS*. 111–123.

Hadi Minooei. 2014. Mechanism Design For Covering Problems. *University of Waterloo, PhD Thesis* (2014).

Amitav Misra. 2013. Responding Before Electric Emergencies. *http://www.afcom.com/digital-library/pub-type/communique/responding-before-electric-emergencies/* (2013).

Roger B. Myerson. 1981. Optimal Auction Design. *Mathematics of Operations Research* 6, 1 (1981), 58–73.

NYISO. 2015a. 2014 Annual Report. *http://www.nyiso.com/public/webdocs/markets_operations/market_data/demand_response/Demand_Response/Reports_to_FERC/2015/20150115 Trsnmttl Ltr Pblc Rprt 2014 Dmnd Rspns Prgrams.pdf* (2015).

NYISO. 2015b. The New York Independent System Operator. *http://www.nyiso.com/public/index.jsp* (2015).

PJM. 2012. Emergency Demand Response (Load Management) Performance Report – 2012/2013. (Dec. 2012).

PJM. 2015. PJM Interconnection. *http://www.pjm.com/* (2015).

Shaolei Ren and Mohanmmad A. Islam. 2014. Colocation Demand Response: Why do I Turn off My Servers?. In *Proc. of USENIX ICAC*. 201–208.

Research and Markets. 2013. Colocation Market - Worldwide Market Forecast and Analysis (2013 - 2018). (2013).

Weijie Shi, Linquan Zhang, Chuan Wu, Zongpeng Li, and Francis Lau. 2014. An Online Auction Framework for Dynamic Resource Provisioning in Cloud Computing. In *Proc. of ACM SIGMETRICS*. 71–83.

Qihang Sun, Chuan Wu, Shaolei Ren, and Zongpeng Li. 2015. Fair Rewarding in Colocation Data Centers: Truthful Mechanism for Emergency Demand Response. In *Proc. of IEEE/ACM IWQoS*. 359 – 368.

Rahul Urgaonkar, Bhuvan Urgaonkar, Michael J. Neely, and Anand Sivasubramaniam. 2011. Optimal Power Cost Management Using Stored Energy in Data Centers. In *Proc. of ACM SIGMET-*

RICS. 221–232.
Vijay Vasudevan, Amar Phanishayee, Hiral Shah, Elie Krevat, David G. Andersen, Gregory R. Ganger, Garth A. Gibson, and Brian Mueller. 2009. Safe and Effective Fine-grained TCP Retransmissions for Datacenter Communication. In *Proc. of ACM SIGCOMM*. 303–314.
William Vickrey. 1961. Counterspeculation, Auctions, and Competitive Sealed Tenders. *The Journal of Finance* 16, 1 (1961), 8–37.
Adam Wierman, Zhenhua Liu, Iris Liu, and Hamed Mohsenian-Rad. 2014. Opportunities and Challenges for Data Center Demand Response. In *Proc. of IEEE IGCC*. 1–10.
Linquan Zhang, Zongpeng Li, and Chuan Wu. 2014a. Dynamic Resource Provisioning in Cloud Computing: A Randomized Auction Approach. In *Proc. of IEEE INFOCOM*. 433–441.
Linquan Zhang, Zongpeng Li, and Chuan Wu. 2014b. Randomized Auction Design for Electricity Markets Between Grids and Microgrids. In *Proc. of ACM SIGMETRICS*. 99–110.
Linquan Zhang, Zongpeng Li, Chuan Wu, and Shaolei Ren. 2015. Online Electricity Cost Saving Algorithms for Co-Location Data Centers. In *Proc. of ACM SIGMETRICS*. 463–464.
Zhi Zhou, Fangming Liu, Zongpeng Li, and Hai Jin. 2015. When Smart Grid Meets Geo-distributed Cloud: An Auction Approach to Datacenter Demand Response. In *Proc. of IEEE INFOCOM*. 2650–2658.

Received October 2015; revised March 2016; accepted May 2016