This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCC.2018.2836424, IEEE Transactions on Cloud Computing

Aggregation-Based Colocation Datacenter Energy Management in Wholesale Markets

Yuanxiong Guo, *Member, IEEE,* Miao Pan, *Member, IEEE,* and Yanmin Gong, *Member, IEEE*

Abstract—In this paper, we study how colocation datacenter energy cost can be effectively reduced in the wholesale electricity market via cooperative power procurement. Intuitively, by aggregating workloads and renewables across a group of tenants in a colocation datacenter, the overall power demand uncertainty of the colocation datacenter can be reduced, resulting in less chance of being penalized when participating in the wholesale electricity market. We use cooperative game theory to model the cooperative electricity procurement process of tenants as a cooperative game, and show the cost saving benefits of aggregation. Then, a cost allocation scheme based on the marginal contribution of each tenant to the total expected cost is proposed to distribute the aggregation benefits among the participating tenants. Besides, we propose proportional cost allocation scheme to distribute the aggregation benefits among the participating tenants after realizations of power demand and market prices. Finally, numerical experiments based on real-world traces are conducted to illustrate the benefits of aggregation compared to noncooperative power procurement.

Index Terms—Cooperative game, colocation datacenter, energy management, wholesale electricity market, cost allocation.

1 Introduction

With the booming of Internet-based and cloud computing services in recent years, datacenters hosting these services have become ubiquitous in every sector of our economy, and their energy consumption has been skyrocketing. According to a report [1] by the Natural Resources Defense Council, datacenters in the U.S. consumed about 91 billion kWh of electricity in 2013, representing 2% of total U.S. electricity consumption and costing U.S. businesses $13 billion in annual electricity bills, and their total electricity consumption is estimated to be 139 billion kWh in 2020. Energy cost accounts for a significant fraction (about 42%) of the datacenter operating expense [2], and this fraction is growing at an alarming rate of 12% annually [3]. Consequentially, reducing energy cost has become a critical concern for datacenter operators.

In order to reduce the growing electricity bills of datacenters, from the demand side, substantial efforts have been made, ranging from hardware such as energy-efficient servers, storage devices, and network switches, to software such as virtualization and dynamic CPU speed scaling and capacity provisioning, which have led to dramatic improvements in the energy-efficiency of datacenters. On the other hand, it is also important for datacenters to manage their energy cost from the supply side. As large consumers, datacenters typically have multiple options to procure electricity to meet their power demand. For instance, a datacenter may purchase power from a retailer such as a local utility company with a pre-specified rate by signing bilateral contracts beforehand, or operate by leveraging on-site power generators and energy storage systems [5].

- *Y. Guo and Y. Gong are with the School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK 74078, USA. E-mail: {richard.guo, yanmin.gong}@okstate.edu.*
- *M. Pan is with the Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77204, USA. E-mail: mpan2@uh.edu.*

Given the significant power consumption and deregulation of electricity market, another promising opportunity to reduce datacenter energy cost is emerging: datacenters can directly participate in the wholesale electricity market to meet their power demand. While it is typical for consumers to buy electricity from local utility companies, some independent system operators (ISOs), such as Electric Reliability Council of Texas (ERCOT) [6] and California ISO [7], have recently developed a market that allows consumers to purchase electricity directly from power suppliers by actively participating in the electricity market. Indeed, datacenter operators like Google have been granted the authority to trade in the wholesale electricity market for the purpose of managing their own energy cost [8]. *The key advantage for datacenters to procure electricity from the wholesale electricity market instead of a local utility company is that they can avoid the insurance premiums, service charges, and mark-up included by utilities in retail rates* [9].

However, a major challenge for datacenters in procuring power directly from the wholesale electricity market is the uncertainty of market prices and their power demand. In most regions of U.S., the wholesale electricity market for electrical power is organized into a two-settlement structure: the day-ahead forward market and the real-time balancing market. The consumers need to make a commitment or bid about their scheduled energy usage to the day-ahead market at first, and then any deviations between the scheduled and actual usage are settled in the real-time balancing market and subject to financial penalties. Since the day-ahead market is often closed several hours (e.g., 14 to 38 hours in California ISO) ahead of the actual operating time, this leaves datacenters vulnerable to high deviation penalties due to their highly uncertain power demand. In addition, market prices are uncertain and hard to predict as well due to the dynamic nature of the market. Therefore, it is imperative for datacenters to mitigate risks associated with these sources of uncertainty in order to maximize the cost

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCC.2018.2836424, IEEE Transactions on Cloud Computing

saving in procuring power from the wholesale electricity market directly.

In this paper, we aim to address the above challenge and optimize datacenter participation strategies in the wholesale electricity market for minimizing energy cost of datacenters. Specifically, we focus on an important but under-explored type of datacenters: colocation datacenters. In a colocation datacenter, multiple tenants house their servers at a shared place. There are more than 1200 colocation datacenters in US and its market is around $43 billion with annual growth rate to be 11% [1]. Although it is risky for tenants to participate in the market individually due to the uncertainty of their workload arrivals and possible on-site renewable generation, this paper takes an aggregation-based approach that transforms these independent tenants from isolated entities into coordinated ones in the market. Our essential idea is to exploit the statistical diversity of workloads and renewables across different tenants and incentivize them to bid collectively in the day-ahead market. Intuitively, by aggregating workloads and renewables from different tenants, the uncertainty of total power demand can be reduced, resulting in less chance of being penalized for deviations in the real-time balancing market and higher energy cost saving.

To incentivize aggregation and distribute aggregation benefits among tenants, we propose to use cooperative game theory. Specifically, the problem can be formulated into a cooperative game with transferrable cost. In this game, the set of players is the set of tenants who seek to cooperate in reducing electricity cost. We first prove that coalitional formation can reduce energy cost compared to individual power procurement in the wholesale electricity market. Then our cooperative game is shown to be balanced and therefore has a nonempty core. Given that the two existing cost allocation methods, the Shapley value and nucleolus, are not applicable to our game, we design an efficient cost allocation scheme that can guarantee mutual benefits for all participating tenants such that no one has the incentive to break up from the coalition and thus locate a cost allocation in the core. Besides, we discuss how to allocate the cost to each tenant after realizations of power demand and market prices. As the cost function of our cooperative game is defined in expectation, there might be some days such that the participating tenants need to pay more compared to the realized cost. Therefore, coalitional members may choose to deviate from such coalition if overpayment keeps occurring. Therefore, we propose a cost allocation method based on the proportion of the realized cost on every day to ensure that in the long run, the allocated realized cost on average will approach the expected cost almost surely.

The rest of the paper is organized as follows. Related work is reviewed in Section 2. A brief overview of cooperative game theory is given in Section 3. In Section 4, we describe the models for tenant power consumption and two-settlement electricity market. In Section 5, we model the datacenter aggregation process as a cooperative game and quantify the benefits of aggregation. Then, the core of the formulated game is shown to be nonempty, and an efficient scheme is proposed to find a cost allocation belonging to the core, and the sharing of realized cost is discussed in Section 6. Simulation results based on real-world traces are presented in Section 7. Finally, the conclusion is given in Section 8.

2 Related Work

In the past decade, multiple schemes have been proposed to reduce the electricity bill of datacenters. From the demand side, in terms of engineering approaches, energy-efficient servers, storage devices and network switches and advanced cooling have been designed to improve the energy efficiency. On the other hand, in terms of algorithmic approaches, dynamic capacity provisioning [10] is developed to reduce energy cost by dynamically turning off surplus servers. Dynamic CPU speed scaling [11] is shown to reduce the energy usage of datacenters by dynamically adapting the processing speed of a server to the current workload. Geographical load balancing [12], [13] is developed to exploit the spatial diversity of electricity prices to minimize the energy cost of geographically distributed datacenters by dynamically routing the user requests to regions with lower energy prices. Exploiting the temporal diversity of electricity prices to reduce energy cost by using energy storage systems or shifting delay-tolerant workload to off-peak time periods has also been investigated in [14], [15], [16], [18]. From the supply side, datacenters can purchase electricity from the retail market with a fixed electricity price by signing bilateral contracts beforehand [17]. On-site renewable power generators such as solar panels and/or wind turbines can also be utilized to reduce energy cost [5], [18], [19]. However, these work does not consider the cost saving opportunities of procuring electricity directly from the wholesale market.

The participation of datacenters in the wholesale electricity market to manage their energy cost has been considered in a few recent studies [4], [20], [21]. However, all of them focus on geo-distributed datacenters with the same owner participating in different wholesale electricity markets and solve the problem using optimization. In contrast to them, in this paper we are the first to consider the colocation datacenter where independent tenants colocated together at the same place jointly participate in the wholesale electricity market. Therefore, we need to use game-theoretic methods to model this multi-agent problem instead of optimization approaches used in geo-distributed datacenters.

3 Background: Cooperative Game Theory

In this section, we will briefly introduce the fundamental concepts of cooperative game theory including the definition for a cooperative game with transferable cost, the solution concept (i.e., the core) of a cooperative game, two types of cooperative games with nonempty core (i.e., the convex games and balanced games), and widely-used cost allocation methods (i.e., the Shapley value and nucleolus).

3.1 Cooperative Game with Transferable Cost

In general, a cooperative game is defined by a pair $(\mathcal{N}, c)$. The first element is the set of players $\mathcal{N} := \{1, 2, \ldots, N\}$, indexed by $i \in \mathcal{N}$. Players may form different coalitions $S \subseteq \mathcal{N}$ to pay a collective cost. The grand coalition $\mathcal{N}$ is the set of all players. Secondly, $c : 2^N \to \mathbb{R}$ is the cost function that assigns a cost to each coalition $S \subseteq \mathcal{N}$. Transferable cost implies that the total cost represented by a real number can be divided in any manner among the coalitional members [22].

3.2 Imputations and the Core

The cost function of a cooperative game is said to be subadditive if it satisfies the following condition:

$$c(\mathcal{S}) + c(\mathcal{T}) \geq c(\mathcal{S} \cup \mathcal{T}),\ \forall \mathcal{S}, \mathcal{T} \subseteq \mathcal{N},\ S \cap T = \emptyset. \tag{1}$$

For such cooperative game, it is to the mutual benefit of the players to form the grand coalition $\mathcal{N}$, since by subadditivity the amount received, $c(\mathcal{N})$, is at least as small as the total amount received by any disjoint set of coalitions they could form. Next, we focus on how to split this amount among participating players.

A cost allocation for the coalition $S \subseteq \mathcal{N}$ is a vector $\boldsymbol{\pi} \in \mathbb{R}^N$ whose entry π_i is the cost dispatched to each player i in the coalition S ($\pi_i = 0, i \notin S$). Further, a cost allocation $\boldsymbol{\pi}$ is said to be *efficient* if $\sum_{i \in \mathcal{N}} \pi_i = c(\mathcal{N})$, i.e., the total amount received by the players should be equal to $c(\mathcal{N})$. A cost allocation $\boldsymbol{\pi}$ is said to be *individually rational* if $\pi_i \leq c(\{i\})$, i.e., no player will be expected to receive more cost than acting individually. A cost allocation $\boldsymbol{\pi}$ for the grand coalition is said to be an *imputation* if it is both efficient and individually rational. In cooperative game theory [23], [24], the set of imputations for the game $(\mathcal{N}, c)$ is defined as

$$\mathcal{I} = \left\{ \boldsymbol{\pi} \in \mathbb{R}^N : \sum_{i \in \mathcal{N}} \pi_i = c(\mathcal{N}),\ \pi_i \leq c(\{i\}),\ \forall i \in \mathcal{N} \right\}. \tag{2}$$

Next, we introduce the solution concept of a cooperative game. The *core* for the game $(\mathcal{N}, c)$ is defined as

$$\mathcal{C} = \left\{ \boldsymbol{\pi} \in \mathbb{R}^N : \sum_{i \in \mathcal{N}} \pi_i = c(\mathcal{N}),\ \sum_{i \in S} \pi_i \leq c(S),\ \forall S \subseteq \mathcal{N} \right\}. \tag{3}$$

The core is a set of imputations such that no coalitions can obtain a cost which is less than the sum of cost assigned by forming the grand coalition. Obviously, if one can locate a cost allocation vector that lies in the core, then the grand coalition is optimal for the cooperative game.

3.3 Convex and Balanced Games

The core is always well-defined, but can be empty. However, the convex games and balanced games are two types of cooperative games which guarantee the existence of *nonempty* core [25], [26]. A cooperative game is said to be convex if the cost function satisfies the following condition:

$$c(\mathcal{S}) + c(\mathcal{T}) \geq c(\mathcal{S} \cup \mathcal{T}) + c(\mathcal{S} \cap \mathcal{T}),\ \forall \mathcal{S}, \mathcal{T} \subseteq \mathcal{N}. \tag{4}$$

This implies the cooperative game has a submodular cost function.

A map $\rho : 2^{\mathcal{N}} \rightarrow [0, 1]$ is said to be balanced if for all $i \in \mathcal{N}$,

$$\sum_{S \in 2^{\mathcal{N}}} \rho(S) \mathbf{1}\{i \in S\} = 1, \tag{5}$$

where $\mathbf{1}\{\cdot\}$ denotes the indicator function. Thus, the balanced map indicates that the sum of weights $\rho(S)$ assigned for each coalition including player i will be equal to 1. Then a cooperative game is said to be balanced if and only if for any balanced map ρ,

$$\sum_{S \in 2^{\mathcal{N}}} \rho(S) c(S) \geq c(\mathcal{N}). \tag{6}$$

3.4 Shapley Value

The Shapley value [27] as the cost allocation method is a unique mapping ψ that satisfies a series of characteristic axioms such as efficiency, symmetry, dummy and additivity. For a cooperative game $(\mathcal{N}, c)$ with transferable cost, the Shapley value $\psi_i(c)$ that distributes the cost for each player $i \in \mathcal{N}$ is defined as

$$\psi_i(c) = \sum_{S \subseteq \mathcal{N} \setminus \{i\}} \frac{|S|!(N - |S| - 1)!}{N!} \left[c(S \cup \{i\}) - c(S) \right]. \tag{7}$$

We observe that in (7), the marginal contribution of each player is represented as $c(S \cup \{i\}) - c(S)$ and the coefficient ahead of the marginal distribution is the probability that the player i randomly joins the coalition S. Thus, the Shapley value can be interpreted as the expected marginal contribution of player i in the grand coalition $\mathcal{N}$ when it joins the coalition S in a random order. It is guaranteed that the Shapley value lies in the core if the game is convex [25].

3.5 Nucleolus

The nucleolus [28] is another common cost allocation method. It uniquely exists in a cooperative game and satisfies the efficiency, individually rational, symmetry and dummy properties [22]. Different from axiomatically designing the cost allocation scheme to ensure fairness as in the Shapley value, the nucleolus aims at minimizing the dissatisfaction of the players. The dissatisfaction of a coalition S given an imputation $\boldsymbol{\pi}$ is measured by the excess. The definition of excess is given by

$$e(\boldsymbol{\pi}, S) = \sum_{i \in S} \pi_i - c(S). \tag{8}$$

Since the core is defined as the set of imputations such that $\sum_{i \in S} \pi_i \leq c(S)$ for all coalitions $S \subseteq \mathcal{N}$, it follows that an imputation $\boldsymbol{\pi}$ is in the core if and only if all its excesses are negative or zero [29]. In order to find the nucleolus, we first need to locate an imputation that minimizes the maximum of the excesses $e(\boldsymbol{\pi}, S)$ over all coalitions S by solving a linear program. After this is done, one may have to solve a second linear programming problem to minimize the next largest excess, and so on. Therefore, in the worst-case, $\mathcal{O}(2^N)$ linear programs need to be solved, which is computationally expensive.

4 System Model

Consider a set $\mathcal{N} := \{1, 2, \ldots, N\}$ of independent tenants in a wholesale colocation datacenter where each tenant pays for their own energy consumption. These tenants may also be equipped with renewable power generators such as solar panels and/or wind turbines. As shown in Fig. 1(a), each tenant can bid its power demand in the wholesale electricity market, and then pay its electricity bill

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCC.2018.2836424, IEEE Transactions on Cloud Computing

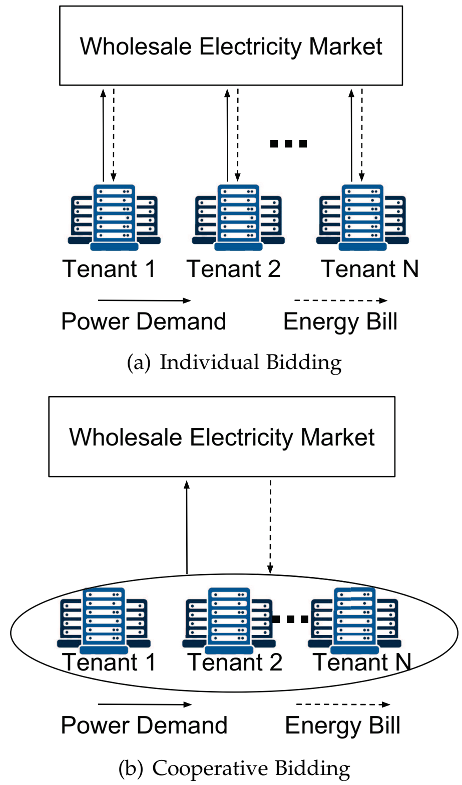


Fig. 1. Individual bidding and cooperative bidding in the wholesale electricity market.

individually. Note that tenants can bid negative amount to supply power in the wholesale electricity market. As shown in Fig. 1(b), we explore the scenario in which individual tenants form a coalition under the coordination of the colo operator to collectively bid their aggregated power demand in the wholesale electricity market as a single entity for cost saving. Without loss of generality, in the following of the paper we restrict our analysis to a specific operating hour.

4.1 Datacenter Power Model

Assume each tenant in the colocation datacenter $i \in \mathcal{N}$ has M_i homogenous servers whose idle and peak power consumption are P_i^{idle} and P_i^{peak}, respectively[1]. Users submit their requests (e.g., search queries) to tenants, and tenants process these requests to satisfy the quality-of-service (QoS) requirement as indicated by the service-level agreement (SLA). When tenant i keeps m_i active servers to process the arriving user requests, its IT power consumption can be estimated as [30]

$$P_i = m_i \left[P_i^{\text{idle}} + u_i (P_i^{\text{peak}} - P_i^{\text{idle}}) \right], \tag{9}$$

where u_i is the average CPU utilization level across all servers at tenant i.

We adopt a M/GI/1 Processor Sharing (PS) queue to model the service process at each server [12]. The workload arrival rate at each tenant i, measured in terms of the average number of arriving user requests per unit time, is assumed to be λ_i, where $\lambda_i \in [\lambda_i^{\min}, \lambda_i^{\max}]$, and $\lambda_i^{\min}$ and $\lambda_i^{\max}$ denotes the minimum and maximum workload arrival rates at each tenant i, respectively. Let μ_i denote the service rate at which user requests are processed by a server at tenant i. Then the average CPU utilization level in tenant i is calculated as $u_i = \lambda_i/(m_i \mu_i)$. Therefore, the power consumption model (9) can be rewritten as

$$P_i = m_i P_i^{\text{idle}} + \frac{\lambda_i}{\mu_i} \left(P_i^{\text{peak}} - P_i^{\text{idle}} \right). \tag{10}$$

Since each user request has a QoS requirement, tenants need to turn on enough servers to meet that requirement. Here we use the average response time as the QoS metric. Based on the M/GI/1/PS queuing model, the average response time of user requests given m_i active servers in tenant i is represented as

$$T_i = \frac{1}{\mu_i - \lambda_i/m_i}. \tag{11}$$

Let $T_i^{\max}$ denote the maximum average response time of user requests that can be tolerated at tenant i. Then to ensure that $T_i \le T_i^{\max}$, we obtain the following feasible range for the number of active servers at tenant i:

$$\frac{\lambda_i}{\mu_i - 1/T_i^{\max}} \le m_i \le M_i. \tag{12}$$

Here, we relax the constraint that requires m_i to be integer given the fact that tenants usually contain thousands of servers. It is assumed that each tenant turn on the minimal number of active servers without violating their QoS requirement using the dynamic capacity provisioning technique [10], [31]. Therefore the IT power consumption of each tenant i is

$$P_i = \frac{\lambda_i}{\mu_i - 1/T_i^{\max}} P_i^{\text{idle}} + \frac{\lambda_i}{\mu_i} \left(P_i^{\text{peak}} - P_i^{\text{idle}} \right). \tag{13}$$

In order to incorporate the non-IT (e.g. cooling, lighting) power consumption of tenants, we denote the average power usage effectiveness (PUE) as γ_i, which is defined as the ratio of the total power consumption to the IT power consumption at tenant i. It follows that the total power consumption E_i of tenant i is given by

$$E_i = \theta_i \lambda_i, \tag{14}$$

where θ_i is a constant defined as

$$\theta_i := \gamma_i \left(\frac{P_i^{\text{idle}}}{\mu_i - 1/T_i^{\max}} + \frac{P_i^{\text{peak}} - P_i^{\text{idle}}}{\mu_i} \right). \tag{15}$$

We have $E_i \in [E_i^{\min}, E_i^{\max}]$, where $E_i^{\min}$ and $E_i^{\max}$ denotes the minimum and maximum power consumption at tenant i, respectively, which depends on the minimum and maximum workload arrival rates $\lambda_i^{\min}$ and $\lambda_i^{\max}$, respectively.

Besides, we assume tenants are equipped with renewable power generators such as solar panels and/or wind turbines, and the renewable power generation is denoted as R_i for each tenant i, where $0 \le R_i \le R_i^{\max}$, and $R_i^{\max}$ is the installed capacity of the renewable power generators at tenant i. Then, the net power demand for each tenant i is given by

$$D_i = E_i - R_i = \theta_i \lambda_i - R_i. \tag{16}$$

1. Note that a tenant with heterogenous servers can be also viewed as several tenants, each having homogeneous servers. Therefore, we focus on the homogenous case in this paper.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCC.2018.2836424, IEEE Transactions on Cloud Computing

We have $D_i \in [D_i^{\min}, D_i^{\max}]$, where $D_i^{\min}$ and $D_i^{\max}$ denote the minimum and maximum net power demand at tenant i, respectively. It follows that $D_i^{\min} = E_i^{\min} - R_i^{\max}$ and $D_i^{\max} = E_i^{\max}$.

When tenant i bids in the day-ahead market one day ahead, the workload arrivals and on-site renewable generation for the next day are uncertain, and thus the workload arrival rate λ_i and renewable power generation R_i can be modeled as random variables whose probability distribution can be empirically estimated from historical data. It follows that the tenant net power demand $D_i(\lambda_i, R_i)$ as a function of the workload arrival rate λ_i and the renewable power generation R_i is also a random variable.

4.2 Two-Settlement Electricity Market

Consider a wholesale electricity market managed by an ISO with a two-settlement structure in the region through which the tenants consume or sell power. It consists of a day-ahead forward market and a real-time balancing market. In the day-ahead forward market, participants bid and schedule power transactions for each hour of the following day before the gate closure. After that, the ISO clears the market and calculates the day-ahead market clearing price for each hour as the intersection between the aggregate supply and demand curves. For instance, for California ISO, the day-ahead forward market closes for bids and schedules by 10 AM and clears by 1 PM on the day prior to the operating day. The schedules cleared in the day-ahead market are financially binding. Any deviations between the day-ahead committed schedule and actual power consumption or supply will be settled in the real-time balancing market during the operating day. If the actual consumption is more than or generation is less than the committed schedule, the energy shortfall will be purchased in the balancing market at the negative imbalance price, which is usually higher than the day-ahead price. If the actual consumption is less than or generation is more than the committed schedule, the energy surplus will be sold at the positive imbalance price, which is usually lower than the day-ahead price. Therefore, power deviations from day-ahead commitments normally result in penalties for participants.

Specifically, for the considered wholesale electricity market, let $p^d \in \mathbb{R}^+$ be the market clearing price in the day-ahead forward market, $p^- \in \mathbb{R}^+$ be the negative imbalance price for energy shortfall, and $p^+ \in \mathbb{R}^+$ be the positive imbalance price for energy surplus. The tenants are assumed to be price-taking because their energy consumption or supply are often too small to influence the market. The market prices (p^d, p^-, p^+) are not known to the tenants at the time of bidding in the day-ahead market and therefore modeled as random variables with known expected values denoted by μ_p^d, μ_p^-, and μ_p^+, respectively, which can be estimated empirically from historical market data. As explained before, without loss of generality, we assume $\mu_p^+ \leq \mu_p^d \leq \mu_p^-$. Moreover, the market prices (p^d, p^-, p^+) are assumed to be statistically independent of the workload arrival rates and renewable power generation $(\lambda_i, R_i, \forall i)$.

Suppose that each tenant $i \in \mathcal{N}$ bids a power consumption or supply amount Q_i in the day-ahead market. Note that in our problem formulation, we focus on a specific operating hour. With the above models and assumptions, it follows that the expected cost of tenant i from participating in the market individually can be calculated as

$$\Phi_i = \mu_p^d Q_i + \mu_p^- \mathbb{E}[(D_i - Q_i)^+] - \mu_p^+ \mathbb{E}[(Q_i - D_i)^+], \quad (17)$$

where $(x)^+ := \max(x, 0)$. Note that there are two cases for tenant i in the market in (17):

- tenant i behaves as a consumer, i.e., $Q_i \geq 0$. $\mu_p^d Q_i$ denotes the day-ahead trading cost, $\mu_p^- \mathbb{E}[(D_i - Q_i)^+]$ denotes the demand shortfall penalty, and $\mu_p^+ \mathbb{E}[(Q_i - D_i)^+]$ denotes the demand surplus profit.
- tenant i behaves as a producer, i.e., $Q_i < 0$. $\mu_p^d Q_i$ denotes the day-ahead trading profit, $\mu_p^- \mathbb{E}[(D_i - Q_i)^+]$ denotes the supply shortfall penalty, and $\mu_p^+ \mathbb{E}[(Q_i - D_i)^+]$ denotes the supply surplus profit.

Note that if $\Phi_i < 0$, then $|\Phi_i|$ represents the expected profit for tenant i when it bids in the market individually.

5 Coalitional Tenant Bidding

In this section, we start by introducing the tenant aggregation model where multiple tenants can form a coalition to bid in the day-ahead market collectively as shown in Fig. 1(b). Then, it can be verified that by bidding net power demand aggregately in the day-ahead market, the total electricity bill can be effectively reduced based on the fact that tenant aggregation can reduce the uncertainty of the total workload arrivals, renewable generation and associated net power demand.

5.1 Tenant Aggregation as a Cooperative Game

Tenants can form different coalitions and bid collectively in the day-ahead market under the coordination of the colo operator. Any coalition $S \subseteq \mathcal{N}$ represents an agreement among the tenantstenant in S to act as a single entity in the market. The aggregated tenant net power demand of a coalition $S \subseteq \mathcal{N}$ is specified by

$$D_S = \sum_{i \in S} D_i. \quad (18)$$

Further, we denote the cumulative distribution function (CDF) of D_S as

$$F_S(e) = \Pr(D_S \leq e). \quad (19)$$

The corresponding quantile function is given by

$$F_S^{-1}(\varepsilon) = \inf \{e \in [D_S^{\min}, D_S^{\max}] : \varepsilon \leq F_S(e)\}, \quad (20)$$

where $D_S^{\min}$ and $D_S^{\max}$ are the minimum and maximum aggregated net power demand for coalition S. Given the minimum and maximum aggregated power consumption and maximum aggregated renewable generation for coalition S denoted as $E_S^{\min}$, $E_S^{\max}$ and $R_S^{\max}$, respectively, it follows that $D_S^{\min} = E_S^{\min} - R_S^{\max}$ and $D_S^{\max} = E_S^{\max}$.

Next, we use cooperative game theory [32] to model this cooperation process as a cooperative game $(\mathcal{N}, c)$ with transferable cost since it is under a multi-agent scenario where each tenant tends to minimize its own net cost. Note that minimizing the negative cost is equivalent to maximize the profit. In our model, the set of tenants $\mathcal{N}$ is the set of players in the cooperative game. Moreover, we assume each tenant always seeks to minimize its own electricity cost, and then the cost function $c(S)$ associated with every coalition $S \subseteq \mathcal{N}$ is represented as its minimum expected energy cost calculated as

$$\Phi_S = \mu_p^d Q_S + \mu_p^- \mathbb{E}[(D_S - Q_S)^+] - \mu_p^+ \mathbb{E}[(Q_S - D_S)^+], \quad (21)$$

$$c(S) = \min_{Q_S \in [D_S^{\min}, D_S^{\max}]} \Phi_S, \quad (22)$$

where Q_S is the bidding amount of any coalition S in the day-ahead market. We assume the market prices for the coalitional bid is the same as that of individual bids. This assumption is acceptable since the tenants are assumed to be relatively small compared to all other prosumers participating in the electricity market so that their operations have little impact on the cleared prices of the day-head market or real-time market [33]. Solving (22) as a newsvendor problem [34], [35], the optimal day-ahead bid and expected cost are given in the following theorem:

***Theorem* 1.** *The optimal day-ahead bid of any coalition S is given by*

$$Q_S^* = F_S^{-1}(\varepsilon^*), \quad \text{where} \quad \varepsilon^* = \frac{\mu_p^- - \mu_p^d}{\mu_p^- - \mu_p^+}. \quad (23)$$

The optimal expected cost is given by

$$c(S) = \mu_p^+ \int_0^{\varepsilon^*} F_S^{-1}(\theta)\, \mathrm{d}\theta + \mu_p^- \int_{\varepsilon^*}^1 F_S^{-1}(\theta)\, \mathrm{d}\theta. \quad (24)$$

Proof: We first rewrite (22) as below:

$$\begin{aligned} c(S) = \min_{Q_S} \mu_p^d Q_S + \mu_p^- \int_{Q_S}^{D_S^{\max}} (u - Q_S) f_S(u)\, \mathrm{d}u \\ - \mu_p^+ \int_{D_S^{\min}}^{Q_S} (Q_S - u) f_S(u)\, \mathrm{d}u, \end{aligned} \quad (25)$$

where $f_S(\cdot)$ is the corresponding probability density function (PDF) of the CDF as defined in (19). Then by applying the first order optimality condition associated with Leibniz integral rule, we have

$$\mu_p^d - \mu_p^-(1 - F_S(Q_S)) - \mu_p^+ F_S(Q_S) = 0, \quad (26)$$

$$Q_S^* = F_S^{-1}(\varepsilon^*), \quad \text{where} \quad \varepsilon^* = \frac{\mu_p^- - \mu_p^d}{\mu_p^- - \mu_p^+}. \quad (27)$$

The optimal expected cost is given by substituting Q_S^* into (25):

$$\begin{aligned} c(S) &= \mu_p^d Q_S^* + \mu_p^- \int_{Q_S^*}^{D_S^{\max}} (u - Q_S^*) f_S(u)\, \mathrm{d}u \\ &\quad - \mu_p^+ \int_{D_S^{\min}}^{Q_S^*} (Q_S^* - u) f_S(u)\, \mathrm{d}u \\ &= \mu_p^d Q_S^* + \mu_p^- \int_{\varepsilon^*}^1 (F_S^{-1}(\theta) - Q_S^*)\, \mathrm{d}\theta \\ &\quad - \mu_p^+ \int_0^{\varepsilon^*} (Q_S^* - F_S^{-1}(\theta))\, \mathrm{d}\theta \\ &= Q_S^* \underbrace{(\mu_p^d - \mu_p^- + \varepsilon^*(\mu_p^- - \mu_p^+))}_{=0} \\ &\quad + \mu_p^+ \int_0^{\varepsilon^*} F_S^{-1}(\theta)\, \mathrm{d}\theta + \mu_p^- \int_{\varepsilon^*}^1 F_S^{-1}(\theta)\, \mathrm{d}\theta. \end{aligned} \quad (28)$$

□

5.2 The Benefits of Aggregation

Intuitively, no group of tenants can do worse by joining a coalition than by acting noncooperatively since aggregation can reduce uncertainty. We will prove this by the following theorem:

***Theorem* 2.** *Given an arbitrary coalition $S \subseteq \mathcal{N}$, let $\{Q_1, Q_2, \ldots, Q_{|S|}\}$ be a set of $|S|$ individual day-ahead bids. For $Q_S = \sum_{i \in S} Q_i$ we have:*

$$\Phi_S(Q_S) \le \sum_{i \in S} \Phi_i(Q_i). \quad (29)$$

Proof: We introduce an ancillary random variable X_i and rewrite (21) in terms of X_i as follows:

$$X_i := D_i - Q_i, \quad (30)$$

$$\begin{aligned} \Phi_S(Q_S) = \mu_p^d Q_S + \mu_p^- \mathbb{E}\big[(\sum_{i \in S} X_i)^+\big] \\ - \mu_p^+ \mathbb{E}\big[(-\sum_{i \in S} X_i)^+\big], \end{aligned} \quad (31)$$

$$\begin{aligned} \sum_{i \in S} \Phi_i(Q_i) = \mu_p^d \sum_{i \in S} Q_i + \mu_p^- \mathbb{E}\big[\sum_{i \in S} (X_i)^+\big] \\ - \mu_p^+ \mathbb{E}\big[\sum_{i \in S} (-X_i)^+\big]. \end{aligned} \quad (32)$$

By adopting the equivalent forms of $(x)^+ := (x)^+ := \max(x, 0) = \frac{x + |x|}{2}$, we have

$$\begin{aligned} &\Phi_S(Q_S) - \sum_{i \in S} \Phi_i(Q_i) = \\ &\mu_p^- \mathbb{E}\left[\frac{\sum_{i \in S} X_i + |\sum_{i \in S} X_i|}{2} - \sum_{i \in S} \frac{X_i + |X_i|}{2}\right] \\ &- \mu_p^+ \mathbb{E}\left[\frac{|\sum_{i \in S} X_i| - \sum_{i \in S} X_i}{2} - \sum_{i \in S} \frac{|X_i| - X_i}{2}\right] \\ &= \left(\frac{\mu_p^- - \mu_p^+}{2}\right) \mathbb{E}\big[(|\sum_{i \in S} X_i| - \sum_{i \in S} |X_i|)\big] \le 0. \end{aligned} \quad (33)$$

The above inequality holds according to the triangle inequality, i.e., $|\sum_{i \in S} X_i| \le \sum_{i \in S} |X_i|$ and also by assumption, we have $\mu_p^- \ge \mu_p^+$. Therefore, $\Phi_S(Q_S) \le \sum_{i \in S} \Phi_i(Q_i)$. □

It is straightforward to see that the expected cost by participating in the market collectively is less than the sum of that by participating in the market individually. That is, the tenants save the expected cost of $\sum_{i \in S} \Phi_i(Q_i) - \Phi_S(Q_S)$ collectively via aggregation. Further, we establish some properties of the cost function associated with every coalition.

***Lemma* 1.** *The optimal expected cost $c(S)$ of any coalition S has following properties:*

1) *Positive homogeneity: For any scalar $\beta \ge 0$, $c(\beta S) = \beta c(S)$.*
2) *Subadditivity: For any two disjoint coalitions S_1 and S_2, if coalition $S_1 \cup S_2$ forms, then $c(S_1 \cup S_2) \le c(S_1) + c(S_2)$.*

Proof: First we prove the positive homogeneity. The CDF of the positively scaled D_S is denoted as

$$F_{\beta S}(u) = \Pr(\beta D_S \le u) = F_{\beta S}\left(\frac{u}{\beta}\right).$$

It follows that the quantile function of $F_{\beta S}(u)$ is given by

$$F_{\beta S}^{-1}(\varepsilon^*) = \beta F_S^{-1}(\varepsilon^*).$$

Using the results from Theorem 1, we can prove the positive homogeneity as

$$\begin{aligned} c(\beta S) &= \mu_p^+ \int_0^{\varepsilon^*} F_{\beta S}^{-1}(\theta)\,\mathrm{d}\theta + \mu_p^- \int_{\varepsilon^*}^1 F_{\beta S}^{-1}(\theta)\,\mathrm{d}\theta \\ &= \beta\left(\mu_p^+ \int_0^{\varepsilon^*} F_S^{-1}(\theta)\,\mathrm{d}\theta + \mu_p^- \int_{\varepsilon^*}^1 F_S^{-1}(\theta)\,\mathrm{d}\theta\right) \\ &= \beta c(S). \end{aligned} \tag{34}$$

Next we prove the subadditivity as

$$\begin{aligned} c(S_1) + c(S_2) &= \min_{Q_{S_1}} \Phi_{S_1}(Q_{S_1}) + \min_{Q_{S_2}} \Phi_{S_2}(Q_{S_2}) \\ &= \Phi_{S_1}(Q_{S_1}^*) + \Phi_{S_2}(Q_{S_2}^*), \end{aligned} \tag{35}$$

where $Q_{S_1}^*$ and $Q_{S_2}^*$ are the optimal day-ahead bids of their respective minimization problems. It follows from Theorem 2 that

$$\begin{aligned} \Phi_{S_1}(Q_{S_1}^*) + \Phi_{S_2}(Q_{S_2}^*) &\geq \Phi_{S_1\cup S_2}(Q_{S_1}^* + Q_{S_2}^*) \\ &\geq \Phi_{S_1\cup S_2}(Q_{S_1\cup S_2}^*) = c(S_1 \cup S_2), \end{aligned}$$

where $Q_{S_1\cup S_2}^*$ is the optimal solution of the expected cost minimization problem under coalition $S_1 \cup S_2$, while $Q_{S_1}^* + Q_{S_1}^*$ is a feasible solution of the minimization problem, then it follows that $c(S_1 \cup S_2) \leq c(S_1) + c(S_2)$. □

From positive homogeneity, we observe that when the aggregated net power demand is scaled, the corresponding value of the optimal expected cost will also be scaled in the same proportion. From subadditivity, we observe that for rational tenants who always try to minimize their cost, they will form a large-size coalition to benefit more from the aggregation. It is straightforward to see in our game that all the tenants will form the grand coalition $\mathcal{N}$ in order to minimize their total expected cost.

6 Cost Allocation Mechanism

In the section, we focus on how to find a cost allocation vector $\boldsymbol{\pi}$ as defined in Section 3.2 to split the total expected cost to each tenant in the grand coalition. First, we verify that our game is nonconvex, and hence the Shapley value is not applicable to locate the core of our game. Next, we show that the core of our cooperative game exists and is nonempty by proving it is a balanced game. Moreover, we propose a cost allocation scheme based on the marginal contribution of each tenant to the total cost in the grand coalition. Last, we discuss how to allocate the cost to each participating tenants after the realizations of net power demand and market prices.

6.1 Existence of the Nonempty Core

As shown in Section 3, both the convexity and balancedness can guarantee the core of a cooperative game to be nonempty. First, we show that our cooperative bidding game is nonconvex by the following theorem:

***Theorem* 3.** *Our cooperative bidding game is nonconvex.*

Proof: We consider a cooperative bidding game involving three tenants, indexed by $i \in \{1,2,3\}$, and denote their net power demand as $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{A}_3$, respectively. We assume the marginal distribution of $\mathcal{A}_1$ and $\mathcal{A}_2$ are given by

$$\mathcal{A}_i = \begin{cases} 2, & \text{w.p. } 0.5 \\ 4, & \text{w.p. } 0.5 \end{cases} \quad \forall i = 1,2.$$

Further, assume $\mathcal{A}_3$ is perfectly positively correlated to $\mathcal{A}_2$, i.e., $\mathcal{A}_3 = \mathcal{A}_2$. We set the expected day-ahead, negative imbalance and positive imbalance prices as $\mu_p^d = 0.9$, $\mu_p^- = 1.4$ and $\mu_p^+ = 0.4$, respectively. Then based on Theorem 1, we have:

$$\begin{aligned} \varepsilon^* &= \frac{1.4-0.9}{1.4-0.4} = 0.5, \\ c(\{1\}) &= c(\{2\}) = c(\{3\}) = 3.2, \\ c(\{1,2\}) &= c(\{1,3\}) = 5.9, \\ c(\{2,3\}) &= 6.4, \\ c(\{1,2,3\}) &= 9.1. \end{aligned}$$

Here, we choose two coalitions as $\mathcal{S} = \{1,2\}$ and $\mathcal{T} = \{1,3\}$, and then from the above example, we have:

$$c(\{1,2\}) + c(\{1,3\}) = 10.8 < c(\{1,2,3\}) + c(\{1\}) = 12.3,$$

which violates the definition of convex game given in (4). Therefore, our cooperative game is nonconvex. □

Since the convexity of a cooperative game is a stronger condition compared to the balancedness, we prove the existence of the core in terms of balancedness by the following theorem:

***Theorem* 4.** *The cooperative game $(\mathcal{N}, c)$ for tenant aggregation is balanced and has a nonempty core.*

Proof: Given an arbitrary balanced map $\rho : 2^{\mathcal{N}} \to [0,1]$, by following the concept of the balanced game, we have

$$\sum_{S\in 2^{\mathcal{N}}} \rho(S)c(S) = \sum_{S\in 2^{\mathcal{N}}} c(\rho(S)S) \tag{36}$$

$$\geq c\left(\sum_{S\in 2^{\mathcal{N}}} \rho(S)S\right) \tag{37}$$

$$\begin{aligned} &= c\left(\sum_{S\in 2^{\mathcal{N}}} \rho(S)\left(\bigcup_{i\in\mathcal{N}} \mathbf{1}\{i\in S\}i\right)\right) \\ &= c\left(\bigcup_{i\in\mathcal{N}}\left(\sum_{S\in 2^{\mathcal{N}}} \rho(S)\mathbf{1}\{i\in S\}\right)i\right) \end{aligned} \tag{38}$$

$$= c\left(\bigcup_{i\in\mathcal{N}} i\right) = c(\mathcal{N}),$$

where (36) is because of the positive homogeneity of $c(S)$, (37) is because of the subadditivity of $c(S)$, and (38) is derived by the definition of balanced map ρ. Therefore, the cooperative game $(\mathcal{N}, c)$ is balanced and has a nonempty core. □

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCC.2018.2836424, IEEE Transactions on Cloud Computing

6.2 Marginal Cost Allocation

Two prominent cost allocation schemes are described in Section 3. However, both of them are not applicable to solve our cooperative game. The Shapley value can be guaranteed to lie in the core if the cooperative game is convex. However, as shown through a counterexample in Theorem 3, our game is not convex. Therefore, the Shapley value does not necessarily belong to the core and hence is not applicable to allocate cost in our game. The nucleolus uniquely exists and can be used as a cost allocation scheme in our game. However, as mentioned before, in the worst-case scenario, $\mathcal{O}(2^N)$ linear programs need to be solved in order to get the cost allocation vector, which is computationally expensive.

Here, we propose a cost allocation scheme based on the marginal contribution of each tenant to the total expected cost when participating in the grand coalition and prove the resulting cost allocation vector is in the core. We define an aggregation level vector $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_N]^T$, where each element $0 \leq \alpha_i \leq 1$ represents the fraction of tenant net power demand D_i that participates in the aggregative power procurement. Thus, the weighted net power demand of the aggregation with the aggregation level vector $\boldsymbol{\alpha}$ is denoted as

$$D_{\boldsymbol{\alpha},\mathcal{N}} = \sum_{i=1}^{N} \alpha_i D_i, \tag{39}$$

whose quantile function is represented by $F^{-1}_{\boldsymbol{\alpha},\mathcal{N}}(\varepsilon)$ and defined similar to (20). Then by applying Theorem 1, we can obtain the optimal expected cost of the weighted net power demand as

$$c_{\boldsymbol{\alpha}}(\mathcal{N}) = \mu_p^+ \int_0^{\varepsilon^*} F^{-1}_{\boldsymbol{\alpha},\mathcal{N}}(\theta)\, \mathrm{d}\theta + \mu_p^- \int_{\varepsilon^*}^{1} F^{-1}_{\boldsymbol{\alpha},\mathcal{N}}(\theta)\, \mathrm{d}\theta. \tag{40}$$

The positive homogeneity and subadditivity proved in Lemma 1 can be easily extended to the case where we consider the weighted optimal expected cost $c_{\boldsymbol{\alpha}}(\mathcal{N})$. Further, we show another property as follows:

***Lemma* 2.** *The weighted optimal expected cost $c_{\boldsymbol{\alpha}}(\mathcal{N})$ of any coalition S is nonincreasing over $\boldsymbol{\alpha}$, i.e., for any two aggregation level vectors, if $\boldsymbol{\alpha} \succeq \boldsymbol{\alpha}'$[2], then $c_{\boldsymbol{\alpha}}(\mathcal{N}) \leq c_{\boldsymbol{\alpha}'}(\mathcal{N})$.*

Proof: Given two aggregation level vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}'$ where $\boldsymbol{\alpha} \succeq \boldsymbol{\alpha}'$, then for any element in the vector $\boldsymbol{\alpha} - \boldsymbol{\alpha}'$, we have $0 \leq \alpha_i - \alpha_i' \leq 1, \forall i \in \mathcal{N}$. Using the subadditivity property, we have

$$c_{\boldsymbol{\alpha}}(\mathcal{N}) \leq c_{\boldsymbol{\alpha}'}(\mathcal{N}) + c_{\boldsymbol{\alpha}-\boldsymbol{\alpha}'}(\mathcal{N}), \tag{41}$$

which indicates the nonincreasing property. □

According to Lemma 2, the optimal expected cost will be achieved when $\boldsymbol{\alpha} = \mathbf{1}$, where $\mathbf{1} \in \mathbb{R}^{N\times 1}$ is an all-one vector. Then it follows that $c_{\boldsymbol{\alpha}}(\mathcal{N})|_{\boldsymbol{\alpha}=\mathbf{1}} = c(\mathcal{N})$.

To distribute the total expected cost $c(\mathcal{N})$ among the tenants in the grand coalition, we compute the expected cost for each tenant i as

$$\pi_i = \frac{\partial c_{\boldsymbol{\alpha}}(\mathcal{N})}{\partial \alpha_i}\Big|_{\boldsymbol{\alpha}=\mathbf{1}}, \ \forall i \in \mathcal{N}. \tag{42}$$

2. The operator $\succeq$ represents component-wise vector comparison.

Indeed, π_i can be decomposed as the multiplication of two terms:

$$\pi_i = \frac{\partial c_{\boldsymbol{\alpha}}(\mathcal{N})}{\partial D_{\boldsymbol{\alpha},\mathcal{N}}}\Big|_{\boldsymbol{\alpha}=\mathbf{1}} \times \frac{\partial D_{\boldsymbol{\alpha},\mathcal{N}}}{\partial \alpha_i}\Big|_{\boldsymbol{\alpha}=\mathbf{1}}, \ \forall i \in \mathcal{N}, \tag{43}$$

where the second term is exactly the net power demand D_i of each tenant. On the other hand, the first term is the partial derivative of the weighted optimal expected cost with respect to the weighted net power demand and then evaluating at the full aggregation level, i.e., $\boldsymbol{\alpha} = \mathbf{1}$, which can be considered as the marginal cost assigned to each tenant. Therefore, the multiplication of the marginal cost and net power demand gives the distributed cost to each tenant. Further, we prove that the cost allocation vector $\boldsymbol{\pi} = [\pi_1, \dots, \pi_N]^T$ given in (42) lies in the core as shown in following theorem:

***Theorem* 5.** *The resulting cost allocation vector of the proposed cost allocation scheme is fair and lies in the core of our cooperative game.*

Proof: Our proof is similar to [36], [37] which focus on different aggregation problems. Here we only give a sketch of the proof process. The basic idea is that we could also use the non-cooperative game theory to model the same problem by allowing power exchange within tenants as well, and our proposed allocation method can find the Nash equilibrium of the formulated noncooperative game. Since the core of our cooperative game can be shown to be the same as the Nash equilibrium of the corresponding noncooperative game, our proposed cost allocation scheme is guaranteed to find the core of the cooperative game. Details about the proof process can be found in [36], [37]. □

The most significant advantage of exploiting this method is its low computational complexity. Compared to using the nucleolus, we only need to calculate $\mathcal{O}(N)$ equations.

6.3 Realized Cost Allocation

Since the cost function (22) of our cooperative bidding game is defined in terms of optimal expected cost, any cost allocation vector lying in the core represents the average cost each participating tenant should pay. However, the realized cost will vary day to day due to the inherent uncertainty of net power demand and market prices. There might be some days such that the participating tenants need to pay more by using our proposed cost allocation method than the realized cost. If overpayment keeps occurring, the coalitional tenants may choose to deviate from the grand coalition, which will break the stability of our game. Therefore, it is necessary to design a way to allocate the realized cost such that the payment to coalition members, averaged over the participating days, approaches the allocated cost in expectation.

Assume the set of operating days $\{1, 2, \dots, K\}$ is indexed by k. After realizations of net power demand of tenants and market prices at a particular hour on day k, we let D_S^k and (p_k^d, p_k^-, p_k^+) denote the aggregated net power demand for coalition S and market prices, respectively. Further, we assume D_S^k and (p_k^d, p_k^-, p_k^+) are independent and identically distributed (i.i.d.) over operating days. Then

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCC.2018.2836424, IEEE Transactions on Cloud Computing

according to (21), we can calculate the realized cost for any coalition $S \subseteq \mathcal{N}$ as

$$\Phi_S^k = p_k^d Q_S^* + p_k^-(D_S^k - Q_S^*)^+ - p_k^+(Q_S^* - D_S^k)^+, \quad (44)$$

where the optimal day-ahead bid Q_S^* is given by Theorem 1.

Then, we denote the realized cost allocation vector at a particular hour on day k as $\boldsymbol{\xi}^k = [\xi_1^k, \ldots, \xi_N^k]^T$ where each entry $\xi_i^k \in \mathbb{R}$ is the realized cost dispatched to tenant i at a particular hour on day k. Given the realization of cost $\Phi_{\mathcal{N}}^k$ for grand coalition $\mathcal{N}$ at a particular hour on day k and the cost allocation vector $\boldsymbol{\pi}^*$ by using our marginal cost allocation method, we propose a proportional allocation to distribute the realized cost to each participating tenant as follows:

$$\xi_i^k = \frac{\pi_i^*}{\sum_{j=1}^N \pi_j^*} \Phi_{\mathcal{N}}^k, \ \forall i \in \mathcal{N}. \quad (45)$$

The above proportional cost allocation method satisfies the following two properties:

- Realized efficiency: $\sum_{i=1}^N \xi_i^k = \Phi_{\mathcal{N}}^k$. The total realized cost at a particular hour on day k paid by all the players should be equal to $\Phi_{\mathcal{N}}^k$. Our proposed method satisfies the realized efficiency since $\sum_{i=1}^N \xi_i^k = \frac{\sum_{i=1}^N \pi_i^*}{\sum_{j=1}^N \pi_j^*} \Phi_{\mathcal{N}}^k = \Phi_{\mathcal{N}}^k$.
- Consistency: $\frac{1}{K}\sum_{k=1}^K \xi_i^k \xrightarrow{a.s.} \pi_i^*$. For player i, the realized cost allocation ξ_i^k at a particular hour averaged over K operating days will approach the expected cost allocation π_i^* almost surely. Our proposed method satisfies the consistency due to the strong law of large numbers since the average of the results obtained from a large number of trials should be close to the expected value.

Due to the above two properties, our proposed proportional cost allocation method can ensure that in the long run, the average of the realized cost allocation will approach the expected cost allocation, which can prevent tenants from leaving the coalition.

7 Numerical Experiments

In this section, we first introduce our simulation setup and then conduct trace-driven simulations to show the benefits of tenant aggregation in trading power in the wholesale electricity market and the effectiveness of our proposed cost allocation scheme.

7.1 Simulation Setup

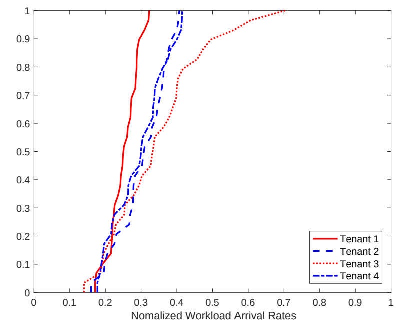
In this following sections, we will introduce the simulation setup for tenants, workloads, renewable energy and electricity prices, respectively. All our simulations are conducted on a desktop computer with an Intel Core i7-4790 3.60GHz CPU and 8GB RAM using MATLAB R2016a.

7.1.1 Colocation Datacenter Descriptions

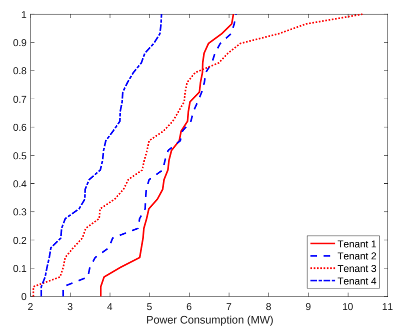
A colocation datacenter with four independent tenants $\mathcal{N} = \{1, 2, 3, 4\}$ is considered in our simulations. The total number of servers for each tenant is 5,000, 7,500, 10,000 and 12,500, respectively. Assume the idle power and peak power of each server is 150 W and 250 W, respectively. Besides, the average PUEs of all the tenants are set to 1.5. The average service rate of a server in each tenant is set to be 200, 250, 300 and 350 requests per second, respectively. The maximum average restenantponse time for each tenant is set to be 100, 80, 60 and 40 ms, respectively. The above simulation parameters are summarized in Table 1.

TABLE 1
Simulation Parameters

	M_i	μ_i (requests/s)	$T_i^{\max}$ (ms)
Tenant 1	5000	200	100
Tenant 2	7500	250	80
Tenant 3	10000	300	60
Tenant 4	12500	350	40



(a)



(b)

Fig. 2. CDFs of the normalized tenant workload arrival rates and power consumption at hour 5.

7.1.2 Workload Descriptions

The real-world dataset we use to simulate the workloads is from the Google cluster trace [38]. The selected dataset includes workload information over 29 days (i.e., 696 hours) during May 2011 for a cluster of 12,500 severs. We repeat the original data and extend it to 1008-hour workloads (i.e., 42 days). Then, we randomly choose 4 different 720-hour (i.e., 30 days) portions from the extended dataset as our tenant workloads. Fig. 2(a) shows the CDFs of the normalized tenant workload arrival rates for four tenants at hour 5. Then we can estimate the power consumption of each tenant according to (14). The CDFs of the power consumption for four tenants at hour 5 are depicted in Fig. 2(b).

7.1.3 Renewable Energy Descriptions

We consider each tenant is equipped with on-site wind turbines. The real-world dataset we use to simulate the wind

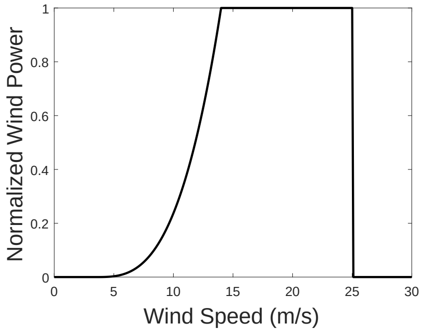


Fig. 3. Normalized wind turbine power curve.

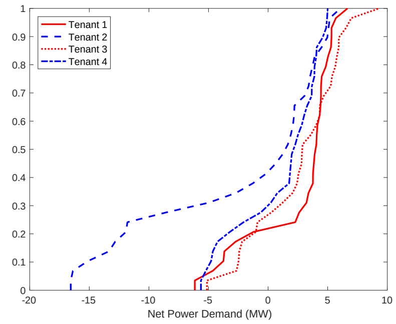


Fig. 5. CDFs of the tenant net power demand at hour 5.

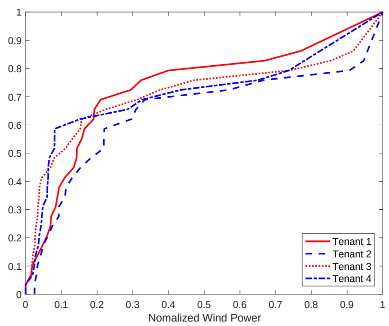


Fig. 4. CDFs of the normalized wind power generation at hour 5.

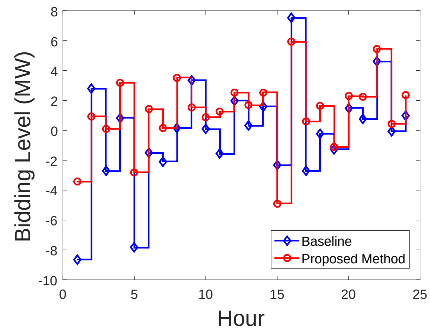


Fig. 6. Day-ahead bidding level comparison over 24 hours.

power generation is from the NREL National Wind Technology Center (M2) [39]. We select the dataset for wind speed at 80 meters from January 2016 to June 2016, and then estimate the corresponding wind power output as shown in Fig. 3, where the cut-in speed, rated output speed and cut-out speed are set to 3.5 m/s, 14 m/s and 25 m/s, respectively. After that, we randomly choose 4 different 720-hour (i.e., 30 days) portions from the converted wind power output data (6 months) as our tenant renewable power generation. The CDFs of the normalized wind power output for four tenants at hour 5 are shown in Fig. 4. Then according to (16), we can obtain the CDFs of net power demand for four tenants at hour 5 which is shown in Fig. 5.

7.1.4 *Electricity Price Descriptions*

In our simulations, tenants can trade power either individually or cooperatively by forming the grand coalition. Moreover, we assume tenants bid their net power demand in the day-ahead market for each hour in the following operating day. By default, the expected day-ahead price μ_p^d is set to be 5 cents/kWh, the expected negative imbalance price μ_p^- is set to be 5.83 cents/kWh, and the expected positive imbalance price μ_p^+ is set to be 2.5 cents/kWh in the simulations.

7.2 Experimental Results

In this section, we simulate and analyze how tenants can benefit from forming the grand coalition to save their electricity cost when trading power in the wholesale electricity market. Here, we consider the case where each tenant bids its net power demand individually by minimizing its expected energy cost as the baseline scenario for comparison.

7.2.1 *Benefits of Aggregation*

We first observe the benefits of coalitional bidding in the wholesale electricity market. Based on Theorem 1, we can calculate the optimal day-ahead bid Q_S^* of any coalition S. Fig. 6 shows the resulting optimal day-ahead bidding level of our proposed method and the sum of optimal individual bidding level in the baseline over 24 hours. It can be observed that the day-ahead bidding level at several hours are negative under baseline scenario, which means at least one tenant behaves as producer by bidding negative power amount in the day-ahead market. Fig. 7 shows the energy cost comparison of our proposed approach and the baseline. The result of the baseline scenario is obtained by adding up the optimal expected electricity cost of each tenant when they bid in the day-ahead market individually, while the result of the proposed method is obtained by letting tenants form the grand coalition to bid in the day-ahead market cooperatively. It is shown in Fig. 7 that the total electricity cost is effectively reduced by cooperative day-ahead bidding, which validates the subadditivity property of our cooperative game given in Lemma 1. The average hourly cost saving is around 18.03% under the current setting.

7.2.2 *Cost Allocation*

Next we focus on how to distribute the total energy cost after coalitional bidding among each participating tenant using our proposed cost allocation method. We split the total expected cost based on the marginal contribution of each tenant in the grand coalition by applying the proposed cost allocation scheme in Section 6.2. Fig. 8 presents the cost allocation to each tenant at hour 5. The height of blue bar and yellow bar denote the individual bidding cost and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCC.2018.2836424, IEEE Transactions on Cloud Computing

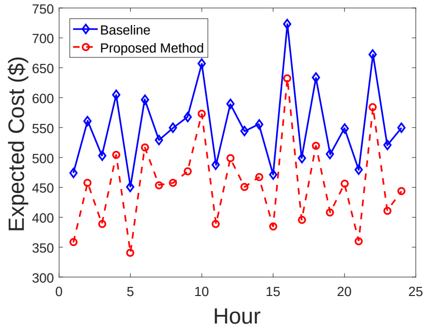


Fig. 7. Total expected cost comparison over 24 hours.

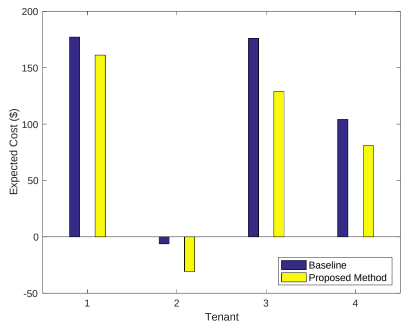


Fig. 8. Cost allocation of each tenant at hour 5 under the current setting.

allocated cost of each tenant after coalitional bidding in the day-ahead market, respectively. It can be observed that tenant 2 behaves as a producer since its individual bidding cost is negative. In order to quantify the aggregation benefits of our proposed method, we define the *cost saving percentage* as the ratio of cost saving and individual bidding cost. The cost saving percentage of each tenant over 24 hours in a day is given in Fig. 9. It can be observed that our proposed allocation method can always ensure positive cost reduction for each tenant and the cost saving amount of each tenant is different, depending on its contribution to the aggregation benefits.

Table 2 presents the noncooperative and coalitional electricity cost of each coalition at hour 5. The last column gives the corresponding excesses $e(\boldsymbol{\pi}, S)$ defined in (8). From row 1 to row 14, the calculated excesses are all negative which satisfies the condition of subgroup rationality, i.e., $\sum_{i\in S}\pi_i \leq c(S)$. The last row indicates that our cost allocation is efficient since $\sum_{i\in\mathcal{N}}\pi_i = c(\mathcal{N})$. It verifies that our proposed cost allocation lies in the core of the cooperative game since both subgroup rationality and efficiency conditions are satisfied.

7.2.3 Impact of Percentile

Now we present how market prices affect the cost saving and the day-ahead bid of each tenant when they form the grand coalition. According to Theorem 1, the optimal day-ahead bid depends on the quantile function where the percentile $\varepsilon^* = \frac{\mu_p^- - \mu_p^d}{\mu_p^- - \mu_p^+}$, which is decided by expected electricity prices μ_p^d, μ_p^- and μ_p^+. In order to obtain different percentiles, we fix the expected day-ahead price μ_p^d and expected positive imbalance price μ_p^+ as constants and adjust the expected negative imbalance price μ_p^- to different values.

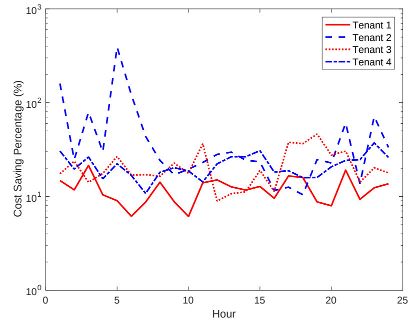


Fig. 9. Individual cost saving percentage of each tenant after coalitional day-ahead bidding over 24 hours.

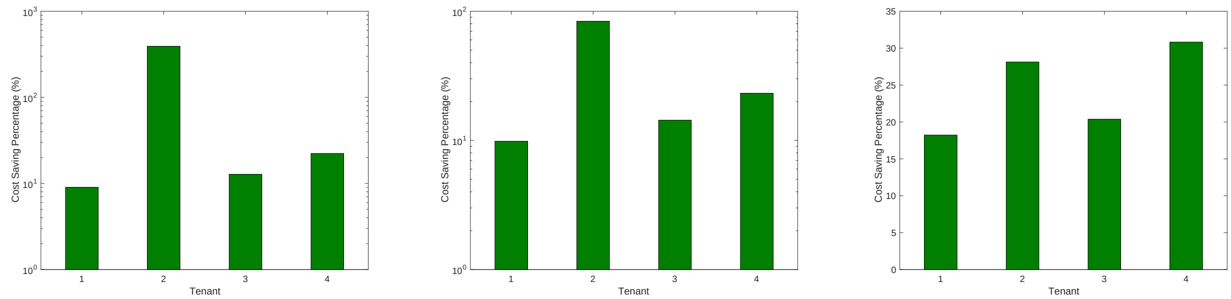
TABLE 2
Cost comparison for all coalitions of four tenants at hour 5

	S	$c(S)$	$\sum_{i\in S}\pi_i$	$\sum_{i\in S}\pi_i - c(S)$
1	$\{1\}$	177.18	161.21	-15.97
2	$\{2\}$	-6.24	-30.72	-24.48
3	$\{3\}$	176.09	128.99	-47.10
4	$\{4\}$	104.21	80.94	-23.27
5	$\{1,2\}$	138.24	130.49	-7.75
6	$\{1,3\}$	321.56	290.20	-31.36
7	$\{1,4\}$	263.08	242.15	-20.93
8	$\{2,3\}$	121.70	98.27	-23.43
9	$\{2,4\}$	56.70	50.22	-6.48
10	$\{3,4\}$	252.68	209.93	-42.75
11	$\{1,2,3\}$	270.45	259.48	-10.97
12	$\{1,2,4\}$	211.28	202.43	-8.85
13	$\{1,3,4\}$	407.32	371.14	-36.18
14	$\{2,3,4\}$	188.49	179.21	-9.28
15	$\{1,2,3,4\}$	340.42	340.42	0

Fig. 10 depicts the cost saving percentage of each tenant at hour 5 when the percentile ε^* is 0.25, 0.50 and 0.75, respectively. Further, the percentage of the average cost saving of each tenant over 24 hours is listed in Table 3. We can observe that for tenant 1, 3 and 4, the percentage of the average cost saving increases when the percentile ε^* increases. This is intuitive since we have less chance to reduce cost through aggregation when the penalty price is lower. Indeed, when the expected negative penalty price is the same as the expected day-ahead electricity price, there is no need for aggregation since one could always buy any shortfall from the real-time market without penalty. However for tenant 2, the percentage of the average cost saving decreases when the percentile ε^* increases. From the net power demand curve of tenant 2, given the percentile ε^*, we can calculate the optimal day-ahead bidding amount. When ε^* increases, the optimal bidding amount changes from negative to positive, and therefore tenant 2 changes from produce to consumer with decreased cost saving percentage.

Fig. 11 shows the changes of day-ahead bidding level of the baseline and the proposed method under different percentiles at hour 5. It can be observed that under both cases, the day-ahead bidding level decreases as the percentile increases. The reason is that when the percentile is near 0, tenants can buy any shortfall in the real-time market without penalty and therefore tend to bid less. On the other

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCC.2018.2836424, IEEE Transactions on Cloud Computing



(a) $\varepsilon^* = 0.25$ (b) $\varepsilon^* = 0.50$ (c) $\varepsilon^* = 0.75$

Fig. 10. Cost saving percentage of each tenant at hour 5 when the percentile ε^* is 0.25, 0.50 and 0.75, respectively.

TABLE 3
The percentage of the average cost saving of each tenant under different percentiles

	$\varepsilon^* = 0.25$	$\varepsilon^* = 0.50$	$\varepsilon^* = 0.75$
Tenant 1	8.90%	12.60%	14.66%
Tenant 2	49.13%	31.35%	23.14%
Tenant 3	10.34%	12.78%	14.82%
Tenant 4	18.63%	22.45%	25.62%

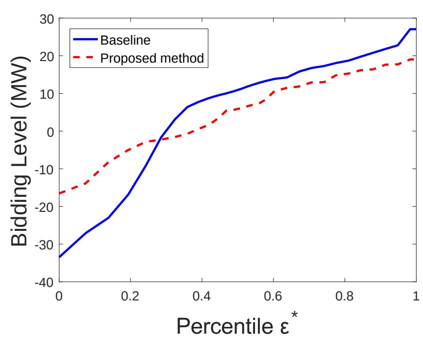


Fig. 11. Day-ahead bidding level comparison under percentiles ε^* from 0 to 1 at hour 5.

hand, when the percentile is approaching 1, tenants behave more conservatively since the expected negative imbalance price is much higher than the expected day-ahead price. In order to avoid high penalty for energy shortfall, they tend to bid more power amount to lower the possible mismatch between committed power supply in the day-ahead market and realized net power demand in the real-time market. Moreover, the change rate of bidding level of our proposed method with respect to the percentile is smaller than that of the baseline. This is due to the fact that the proposed method has a smaller net power demand uncertainty and therefore is less sensitive to the percentile.

8 Conclusion

In this paper, we have proposed a new approach to minimize the electricity cost for tenants in colocation datacenters participating in the wholesale electricity market. The electricity cost can be effectively reduced by bidding in the day-ahead market collectively since aggregation can reduce the uncertainty of net power demand. We model this aggregation process as a cooperative game and present a cost allocation mechanism based on the marginal contribution of each tenant to the total expected cost to distribute the optimal expected cost to each tenant within the grand coalition. Moveover, we have discussed how to share the coalitional cost after the realizations of net power demand and market prices. Our proposed proportional cost allocation method can ensure the stability of our cooperative bidding game after realizations in the long run. Finally, simulations based on real-world traces verify the effectiveness of our proposed cost saving method.

Acknowledgments

This work of M. Pan was partially supported by the U.S. National Science Foundation under grants CNS-1343361, CNS-1350230 (CAREER), CNS-1646607, and CNS-1702850.

References

[1] National Resources Defense Council, “Scaling up energy efficiency across the data center industry: evaluating key drivers and barriers,” 2014.

[2] J. Hamilton, “Cost of power in large-scale data centers,” 2008. [Online]. Available: http://perspectives.mvdirona.com/2008/11/cost-of-power-in-large-scale-data-centers/

[3] R. E. Brown, R. Brown, E. Masanet, B. Nordman, B. Tschudi, A. Shehabi, J. Stanley, J. Koomey, D. Sartor, P. Chan *et al.*, “Report to congress on server and data center energy efficiency: Public law 109-431,” Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, CA (US), Tech. Rep., 2007.

[4] L. Yu, T. Jiang, Y. Cao, and Q. Zhang, “Risk-constrained operation for internet data centers in deregulated electricity markets,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 5, pp. 1306–1316, 2014.

[5] Y. Guo, Y. Gong, Y. Fang, P. P. Khargonekar, and X. Geng, “Energy and network aware workload management for sustainable data centers with thermal storage,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 8, pp. 2030–2042, 2014.

[6] “Electric Reliability Council of Texas.” [Online]. Available: http://www.ercot.com/

[7] “California ISO.” [Online]. Available: http://www.caiso.com/

[8] “Google gets go-ahead to buy, sell energy,” Feb. 2010. [Online]. Available: https://www.cnet.com/news/google-gets-go-ahead-to-buy-sell-energy/

[9] M. Ghamkhari, A. Wierman, and H. Mohsenian-Rad, “Energy portfolio optimization of data centers,” *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1898–1910, 2017.

[10] M. Lin, A. Wierman, L. L. Andrew, and E. Thereska, “Dynamic right-sizing for power-proportional data centers,” *IEEE/ACM Transactions on Networking*, vol. 21, no. 5, pp. 1378–1391, 2013.

[11] A. Wierman, L. L. Andrew, and A. Tang, “Power-aware speed scaling in processor sharing systems,” in *Proc. IEEE INFOCOM*, 2009.

[12] Z. Liu, M. Lin, A. Wierman, S. H. Low, and L. L. Andrew, “Greening geographical load balancing,” in *Proc. ACM SIGMETRICS*, 2011.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCC.2018.2836424, IEEE Transactions on Cloud Computing

[13] A. Qureshi, R. Weber, H. Balakrishnan, J. Guttag, and B. Maggs, "Cutting the electric bill for internet-scale systems," in *Proc. ACM SIGCOMM*, 2009.

[14] Y. Guo, Z. Ding, Y. Fang, and D. Wu, "Cutting down electricity cost in internet data centers by using energy storage," in *Proc. IEEE GLOBECOM*, 2011.

[15] Y. Guo and Y. Fang, "Electricity cost saving strategy in data centers by using energy storage," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 6, pp. 1149–1160, 2013.

[16] Y. Guo, Y. Fang, and P. P. Khargonekar, "Optimal workload and energy storage management for cloud data centers," in *Proc. MILCOM*, 2013.

[17] C. Wang, B. Urgaonkar, G. Kesidis, U. V. Shanbhag, and Q. Wang, "A case for virtualizing the electric utility in cloud data centers." in *HotCloud*, 2014.

[18] Y. Guo, M. Pan, Y. Gong, and Y. Fang, "Dynamic multi-tenant coordination for sustainable colocation data centers," *IEEE Transactions on Cloud Computing*, 2017.

[19] Y. Guo, Y. Gong, Y. Fang, P. P. Khargonekar, and X. Geng, "Optimal power and workload management for green data centers with thermal storage," in *Proc. IEEE GLOBECOM*, 2013.

[20] Y. Zhang, L. Deng, M. Chen, and P. Wang, "Joint bidding and geographic load balancing for datacenters: Is uncertainty a blessing or a curse?" in *Proc. IEEE INFOCOM*, 2017.

[21] T. N. Le, J. Liang, Z. Liu, R. K. Sitaraman, J. Nair, and B. J. Choi, "Optimal energy procurement for geo-distributed data centers in multi-timescale electricity markets," in *Proc. IFIP Performance*, 2017.

[22] W. Saad, Z. Han, M. Debbah, A. Hjorungnes, and T. Basar, "Coalitional game theory for communication networks," *IEEE Signal Processing Magazine*, vol. 26, no. 5, pp. 77–97, 2009.

[23] H. P. Young, *Cost allocation: methods, principles, applications*. North Holland Publishing Co., 1985.

[24] S. H. Tijs and T. S. Driessen, "Game theory and cost allocation problems," *Management Science*, vol. 32, no. 8, pp. 1015–1028, 1986.

[25] L. S. Shapley, "Cores of convex games," *International journal of game theory*, vol. 1, no. 1, pp. 11–26, 1971.

[26] ——, "On balanced sets and cores," *Naval research logistics quarterly*, vol. 14, no. 4, pp. 453–460, 1967.

[27] R. B. Myerson, *Game theory: analysis of conflict*. Harvard University Press, 2007.

[28] G. Owen, *Game Theory*, 4th ed. Emerald Group Publishing Limited, 2013.

[29] T. S. Driessen, *Cooperative games, solutions and applications*. Springer Netherlands, 1988, vol. 3.

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

[31] Q. Wu, "Making Facebook's software infrastructure more energy efficient with Autoscale." [Online]. Available: https://code.facebook.com/posts/816473015039157/

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

[33] M. Kohansal and H. Mohsenian-Rad, "Price-maker economic bidding in two-settlement pool-based markets: The case of time-shiftable loads," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 695–705, 2016.

[34] E. Y. Bitar, R. Rajagopal, P. P. Khargonekar, K. Poolla, and P. Varaiya, "Bringing wind energy to market," *IEEE Transactions on Power Systems*, vol. 27, no. 3, pp. 1225–1235, 2012.

[35] E. Baeyens, E. Y. Bitar, P. P. Khargonekar, and K. Poolla, "Coalitional aggregation of wind power," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 3774–3784, 2013.

[36] Y. Zhao, J. Qin, R. Rajagopal, A. Goldsmith, and H. V. Poor, "Wind aggregation via risky power markets," *IEEE Transactions on Power Systems*, vol. 30, no. 3, pp. 1571–1581, 2015.

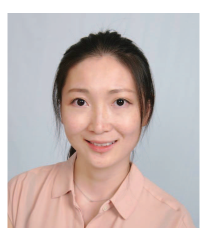
[37] D. Kalathil, C. Wu, K. Poolla, and P. Varaiya, "The sharing economy for the smart grid," *arXiv preprint arXiv:1608.06990*, 2016.

[38] C. Reiss, J. Wilkes, and J. L. Hellerstein, "Google cluster-usage traces: format+ schema," *Google Inc., White Paper*, pp. 1–14, 2011.

[39] "Measurement and instrumentation data center (midc)," http://www.nrel.gov/midc/.

Yuanxiong Guo (M'14) received the B.Eng. degree in electronics and information engineering from Huazhong University of Science and Technology, Wuhan, China, in 2009, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Florida, Gainesville, FL, USA, in 2012 and 2014, respectively. Since 2014, he has been an Assistant Professor with the School of Electrical and Computer Engineering at Oklahoma State University, Stillwater, OK, USA. His current research interests include resource management and cybersecurity for networked systems including cyber-physical systems, Internet of things, wireless networks, and cloud/edge networks. He is a recipient of the Best Paper Award in the IEEE Global Communications Conference 2011. He is a member of IEEE and ACM.

Miao Pan (M'12) received his BSc degree in Electrical Engineering from Dalian University of Technology, China, in 2004, MASc degree in electrical and computer engineering from Beijing University of Posts and Telecommunications, China, in 2007 and Ph.D. degree in Electrical and Computer Engineering from the University of Florida in 2012, respectively. He is now an Assistant Professor in the Department of Electrical and Computer Engineering at University of Houston. He was an Assistant Professor in the Computer Science at Texas Southern University from 2012 to 2015. His research interests include cognitive radio networks, cybersecurity, and cyber-physical systems. His work on cognitive radio network won Best Paper Award in Globecom 2015. Dr. Pan is currently Associate Editor for IEEE Internet of Things (IoT) Journal. He is a member of IEEE.



Yanmin Gong (M'16) received the B.Eng. degree in electronics and information engineering from Huazhong University of Science and Technology, China, in 2009, the M.S. degree in electrical engineering from Tsinghua University, China, in 2012, and the Ph.D. degree in electrical and computer engineering from the University of Florida, Gainesville, FL, USA, in 2016. She has been an Assistant Professor in the School of Electrical and Computer Engineering at Oklahoma State University since August 2016. Her research interests include information security and privacy, mobile and wireless security and privacy, security in Internet-of-Things, and privacy-preserving big data analytics. She has served as the Technical Program Committee members for several conferences including IEEE INFOCOM and CNS. She is a member of IEEE and ACM.