Optimal Pricing Strategy for Data Center Considering Demand Response and Renewable Energy Source Accommodation

Chenwei Jiang, Chung-Li Tseng, Yizheng Wang, Zhou Lan, Fushuan Wen, *Fellow*, *IEEE*, Fei Chen, and Liang Liang

***Abstract*—With the continuous development of information technology, data centers (DCs) consume significant and ever-growing amounts of electrical energy. Renewable energy sources (RESs) can act as clean solutions to meet this requirement without polluting the environment. Each DC serves numerous users for their data service demands, which are regarded as flexible loads. In this paper, the willingness to pay and time sensitivities of DC users are firstly explored, and the user-side demand response is then devised to improve the overall benefits of DC operation. Then, a Stackelberg game between a DC and its users is proposed. The upper-level model aims to maximize the profit of the DC, in which the time-varying pricing of data services is optimized, and the lower-level model addresses user's optimal decisions for using data services while balancing their time and cost requirements. The original bi-level optimization problem is then transformed into a single-level problem using the Karush-Kuhn-Tucker optimality conditions and strong duality theory, which enables the problem to be solved efficiently. Finally, case studies are conducted to demonstrate the feasibility and effectiveness of the proposed method, as well as the effects of the time-varying data service price mechanism on the RES accommodation.**

***Index Terms*—Data center (DC), demand response (DR), pricing strategy, renewable energy source (RES), Stackelberg game.**

Manuscript received: February 19, 2021; revised: July 13, 2021; accepted: August 28, 2021. Date of CrossCheck: August 28, 2021. Date of online publication: November 30, 2021.

This work was supported in part by National Natural Science Foundation of China (No. U1910216) and in part by Science and Technology Project of State Grid Zhejiang Electric Power Co., Ltd. (No. 5211JY19000T).

C. Jiang and Y. Wang are with the School of Electrical Engineering, Zhejiang University, Hangzhou 310027, China, and Y. Wang is also with the Economic and Technology Research Institute, State Grid Zhejiang Electric Power Co., Ltd., Hangzhou 310000, China (e-mail: chenweijiang@zju. edu. cn; yizhengwang@zju.edu.cn).

C.-L. Tseng is with UNSW Business School, The University of New South Wales, Sydney, Australia (e-mail: c.tseng@unsw.edu.au).

Z. Lan is with the Economic and Technology Research Institute, State Grid Zhejiang Electric Power Co., Ltd., Hangzhou 310000, China (e-mail: lanzhou_zju@163.com).

F. Wen (corresponding author) is with the Department of Electrical Power Engineering and Mechatronics, Tallinn University of Technology, Estonia, and he is also with the College of Electrical Engineering, Zhejiang University, Hangzhou, China (e-mail: fushuan.wen@gmail.com).

F. Chen is with the Economic and Technology Research Institute, State Grid Zhejiang Electric Power Co., Ltd., Hangzhou 310000, China (e-mail: fei_ei_chen@zj.sgcc.com.cn).

L. Liang is with the Jiaxing Power Supply Company, State Grid Zhejiang Electric Power Co., Ltd., Jiaxing 314000, China (e-mail: 18805739125@139.com).

DOI: 10.35833/MPCE.2021.000130

I. Introduction

WITH the rapid technological development of renewable energy power generation, economic and efficient means of consuming renewable energy sources (RESs) have become the focuses of both academy and industry [1], [2]. Among many possible solutions, user-side demand response (DR), which has received widespread attention [3]-[5], is an effective method because of its low cost and high flexibility. Some typical DR mechanisms have been proposed in [6], [7] to adjust the users' energy demands and enhance the capability of RES accommodation, which guide users in responding to prices or incentive signals [8] and create effective interactions between the grid and users.

With recent advances in information technology and the development of the Internet, major technology companies such as Google, Microsoft, and Facebook have built their own data centers (DCs) to provide various cloud storage and computing services [9]. As the scale of DCs expands, the energy consumption has grown rapidly. According to reports, China's DC energy consumption in 2017 was 122.15 billion kWh, which exceeded the annual power generation of the Three Gorges Dam during that year [10]. In addition, it is estimated that DCs will account for the largest share of global energy consumption by 2025 by as much as 33% [11]. However, a DC can act as a new type of large-capacity DR resource [12] when real-time response capabilities and the flexible scheduling characteristics of its loads are considered.

Existing research works on DCs and their DR practices have mainly focused on the internal management and load distribution strategies of DCs [13]-[16], the load transfers of geo-distributed DCs [16]-[18], the optimal scheduling of data services that can be delayed [19], [20], and other aspects. An efficient resource management policy for virtualized cloud DCs, which results in substantial energy savings by the dynamic reallocation of virtual machines, is proposed in [13]. A cooling-efficiency-enabled demand-management solution is proposed in [16], in which a DR management model and the virtual-machine live-migration technology are used to reduce the electricity costs in DCs. Reference [17] optimizes a pricing scheme using a two-stage Stackelberg game with the workload leveled not only over time but also over space by transferring loads between geo-distributed DCs.

Reference [19] proposes an approach that divides users' data service demands into real-time delay-sensitive jobs and deferrable but deadline-oriented jobs, and optimizes the scheduling of the latter to reduce DC operation costs.

In general, existing research works on DC participation in DR have mainly focused on optimization within the DC and collaborative scheduling among several DCs. However, relatively few have been conducted on the DR potential of users subordinated to DCs. In addition, existing research works on the user side have usually focused on the optimal scheduling for data service demands of delay-tolerant users (e.g., batch workload) while ignoring the DR potential for delay-sensitive users (e.g., interactive workload). Moreover, compared with the practice of electricity retailers in setting prices [21], the characteristics of data service and electricity demands are different. Thus, it is necessary to establish a data service price mechanism and a DR model for DC users. Therefore, based on the modeling and analysis of the data service demand of DC users, this paper establishes a time-varying DC data service price mechanism that is different from most studies that set the DC data service price to be a constant value. A Stackelberg game is proposed in which the DC offers users a time-varying price of data services to reflect its variable energy costs and incentivize users to participate in DR. The major contributions of this paper are as follows.

1) We propose a model describing a user's willingness to pay (WTP) and establish a time sensitivity, by which a time-varying data service pricing mechanism considering users' loss is then proposed. The data service price is set to reflect the changes in the electricity price in the grid and the purchase price of RES, thus encouraging price-sensitive users to optimize the plans to obtain data services within their adjustable time period.

2) We construct a Stackelberg game between a DC and users. At the upper and lower levels, the goals are to maximize the profit of the DC and to minimize the users' data service cost, respectively. After the nonlinear terms and 0/1 variables in the proposed model are processed, the bi-level optimization problem is transformed into a single-level one using Karush-Kuhn-Tucker (KKT) optimality conditions and strong duality theory.

3) The effects of the price mechanism, which plays a vital role in both the reduction of DC energy cost and the increase in the consumption of RES generation, are fully analyzed. A sensitivity analysis is conducted to show the effects of the proposed mechanism with different quantities of price levels.

The remainder of this paper is organized as follows. The modeling of DC and users' DR are presented in Section II. The optimal pricing strategy and solution procedure are described in Section III. Section IV presents case studies and simulation results. And finally, conclusions are drawn in Section V.

II. Modeling of DC and Users' DR

A. DC Operation Process

The DC operation process and its DR-enabled users are presented in Fig. 1. The energy layer includes RESs and the grid, which provide power supply and electricity price to the DC layer. The DC layer mainly includes a dispatch center and servers under control, where the dispatch center assigns tasks to the servers according to data service demand and provides DR users with the information of day-ahead data service price. The DR user layer includes all users participating in the DR program. Based on price information, DR users can determine their data service time in a more flexible manner to balance their cost requirements.

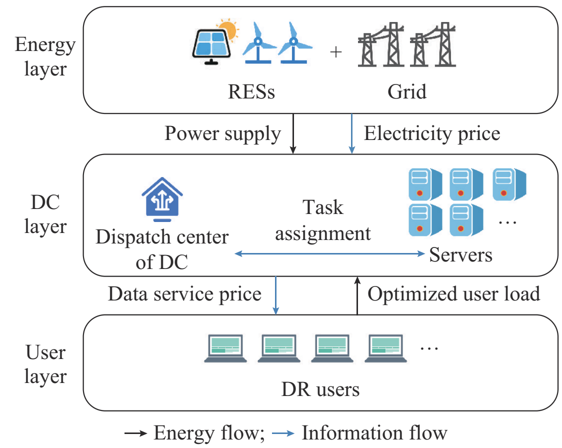

Fig. 1. DC operation process and its DR users.

B. Modeling of DC

1) Energy Consumption of DC

The energy consumption of a DC generally consists of servers, communication and storage, air conditioning, and other components. In general, servers are flexible in shifting loads to accommodate price signals from the energy layer. The power usage efficiency (PUE) is used to estimate the energy consumption of a DC [17].

$$P_t^{\mathrm{DC}} = n_t\left[P^{\mathrm{idle}} + \left(P^{\mathrm{peak}} - P^{\mathrm{idle}}\right)u_t + \left(\eta - 1\right)P^{\mathrm{peak}}\right] \tag{1}$$

where P_t^{DC} is the energy consumption of the DC at time t; P^{peak} and P^{idle} are the peak power and idle power of a server, respectively; $u_t = D_t/(n_t\mu)$ is the average server utilization, D_t is the total incoming workload at time t, n_t is the number of active servers at time t, μ is the service rate of a server; and η is the PUE of the DC. Equation (1) can be rewritten as:

$$P_t^{\mathrm{DC}} = n_t P^{\mathrm{idle}} + \left(P^{\mathrm{peak}} - P^{\mathrm{idle}}\right)D_t/\mu + n_t\left(\eta - 1\right)P^{\mathrm{peak}} \tag{2}$$

2) Quality of Service (QoS)

It is important for DCs to set a maximum response time and provide QoS guarantees to users, which can be indicated in the service level agreement (SLA) [22]. In this paper, the M/M/1 queuing theory [23] is used to analyze the average response time for interactive workload at a DC, where the delay constraints are given by:

$$0 \le \frac{1}{\mu - D_t/n_t} \le d^{\max} \tag{3}$$

$$0 \le n_t \le n^{\max} \tag{4}$$

where $d^{\max}$ is the maximum delay time that a request can tolerate; and $n^{\max}$ is the maximum number of servers in the DC.

3) Server Operation Constraints

The dispatch center of the DC is responsible for scheduling the servers to capitalize on the flexibility of the DC in terms of workload shifting while satisfying the reliability requirement. As a result, a server cannot be switched on or off too frequently [16]. Therefore, it is necessary to establish a minimum on-off time constraint on the servers.

$$\sum_{n=1}^{n^{\max}} X_{n,t}^{\text{server}} = n_t \quad \forall t \in [1, T] \tag{5}$$

$$K_{n,t+1}^{\text{server}} = X_{n,t+1}^{\text{server}} - X_{n,t}^{\text{server}} \quad \forall t \in [1, T-1] \tag{6}$$

$$SX_{n,t_0} = \sum_{t=t_0+1}^{t_0+T^{\mathrm{d}}} X_{n,t}^{\text{server}} \quad \forall t_0 \in [1, T-T^{\mathrm{d}}] \tag{7}$$

$$T^{\mathrm{d}} - \left(1 - K_{n,t+1}^{\text{server}}\right)M \le SX_{n,t} \le T^{\mathrm{d}} + \left(1 - K_{n,t+1}^{\text{server}}\right)M \quad \forall t \in [1, T-T^{\mathrm{d}}] \tag{8}$$

$$0 - \left(1 + K_{n,t+1}^{\text{server}}\right)M \le SX_{n,t} \le 0 + \left(1 + K_{n,t+1}^{\text{server}}\right)M \quad \forall t \in [1, T-T^{\mathrm{d}}] \tag{9}$$

where $X_{n,t}^{\text{server}}$ is a binary variable that describes the start/stop status of server n at time t during T time periods; $K_{n,t}^{\text{server}}$ takes the values of -1, 0, and 1 to describe the change in the status of server n at time t; T^{d} is the shortest interval before a server can be stopped or started; $SX_{n,t}$ is the sum of the statuses of server n for a duration of T^{d} starting from time $t+1$; and M is a sufficiently large positive number.

When server n is started at time t, i.e., $K_{n,t}^{\text{server}}=1$, this server cannot be shut down within the duration of T^{d}, as imposed by (8). Similarly, when server n is shut down at time t, i.e., $K_{n,t}^{\text{server}}=-1$, the server cannot be started within the duration of T^{d}, as imposed by (9).

C. Modeling of User's DR Characteristics

1) Analysis of DR Behavior

All DC users have different WTPs for the service and different time sensitivities for task delays. These two characteristics are critical to the success of a DR program. As shown in Figs. 2 and 3, given a large number of DC users, it is assumed that the WTP and time sensitivity (in terms of delay tolerance) of users follow normal distributions, as denoted by $N^p(\mu^p,(\sigma^p)^2)$ and $N^t(\mu^t,(\sigma^t)^2)$, where μ^p, σ^p, μ^t, and σ^t are the expectations and standard deviations of users' WTP and delay tolerance distributions, respectively.

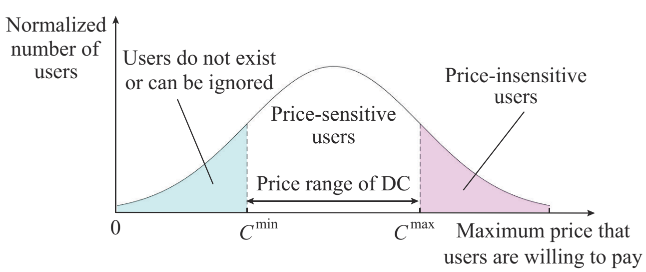

Fig. 2. Users' WTP distribution.

Note that the two distributions shown in Figs. 2 and 3 are plotted with respect to the price and delay policies of the DC. When a users' WTP for the data service is higher than the price range set by the DC, this user can be regarded as a price-insensitive user. Likewise, when a user's delay tolerance is lower than the range specified by the DC, this user is regarded as a time-sensitive user. The other extreme sides of these two distributions are the users who may not afford the service and have great tolerance for time delay, and they are not ideal users for the DR program. Instead, the DR program mainly targets the users who sit in the middle of both distributions, i. e., both price-sensitive and time-insensitive; these users are willing to save some money in exchange for some task delays. Although the sensitivities on both price and time matter, the time insensitivities are more important than these users' WTP for the health of the DR program, as the DR program relies on users' time insensitivities to shift loads for better outcomes.

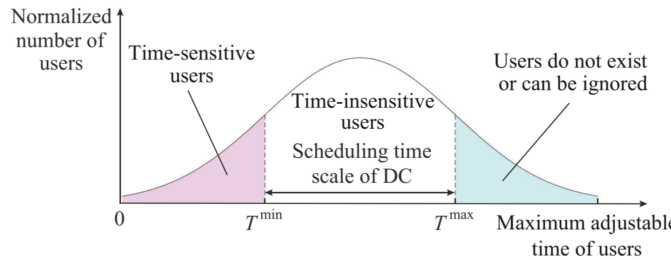

Fig. 3. Users' delay tolerance distribution.

It should be mentioned that interactive workloads for DC users are considered. In other words, the DC must respond within a short time period after a data service demand is lodged. This situation is more realistic. For this type of users, the DC cannot use the time delay strategy for scheduling as proposed in many existing research works. Instead, a time-varying price mechanism is more suitable for this type of service, which will be discussed in detail in Section III.

2) Modeling of DR Pricing Considering Users' DR Characteristics

Let us assume that the DC announces the data service price C_t^{data} (e.g., hourly) for the next day. Here, C_t^{data} is determined by the optimization discussed in Section III. It is also assumed that the WTP of user j is $C_j^{\max}$, with an initial data service time at t_j^{initial}. The following pricing rule is applied to the users based on their DR intentions.

Let the price offered to user j be denoted by C_j^{deal}. If user j is not willing to participate in the DR program, the time of the user's data service task will not be changed, and the fee will be charged at the price corresponding to the original time, i.e.,

$$C_j^{\text{deal}} = C_t^{\text{data}} \tag{10}$$

$$t = t_j^{\text{initial}} \tag{11}$$

These users are time-sensitive. However, for those participating in the DR program, their data service may be changed from t_j^{initial} to sometime earlier or later with the time change capped by $t_j^{\max}$. The service rate is the minimum price during a possible period, i.e.,

$$C_j^{\text{deal}} = \min\left\{ C_t^{\text{data}} \middle| t_j^{\text{initial}} - t_j^{\max} \le t \le t_j^{\text{initial}} + t_j^{\max} \right\} \tag{12}$$

Regardless of the time sensitivity, the WTP of these users is affected by the price of data services. Therefore, they can be called price-sensitive users.

Any user whose WTP is lower than all service rates during the possible period, i.e., $C_j^{\max} < \min\left\{C_t^{\text{data}} \middle| t_j^{\text{initial}} - t_j^{\max} \le t \le t_j^{\text{initial}} + t_j^{\max}\right\}$, is considered to be lost.

3) Modeling of Lost Users

As indicated previously, some users may leave without using the data service if their WTP is too low. Therefore, given the distributions of user WTP and delay tolerance as shown in Figs. 2 and 3, respectively, the lost users must be taken into account. The actual user's demand of the DC at time t can be expressed as:

$$N_t^{\text{final}} = N_t^{\text{basic}} - N_t^{A_1} + N_t^{A_2} - N_t^{A_0} \tag{13}$$

$$N^{\text{user}} = \sum_{t=1}^{T} N_t^{\text{final}} \tag{14}$$

where N_t^{final} is the number of user demands; N_t^{basic} is the number of users who do not participate in the DR program; $N_t^{A_1}$ is the number of time-insensitive users transferred to other time periods from time t; $N_t^{A_2}$ is the number of time-insensitive users transferred from other time periods to time t; $N_t^{A_0}$ is the number of lost users; and N^{user} is the total number of users during T time periods.

III. Optimal Pricing Strategy and Solution Procedure

Figure 4 presents the scheme of the proposed optimal pricing problem of the DC modeled as a Stackelberg game. In other words, the DC anticipates and accounts for the optimal reactions from its users in its optimization. The goal of the upper-level problem is to maximize the revenue by setting optimal data service price schedules to induce DR behaviors, and the goal of the lower-level problem is to minimize the data service cost of the users through DR.

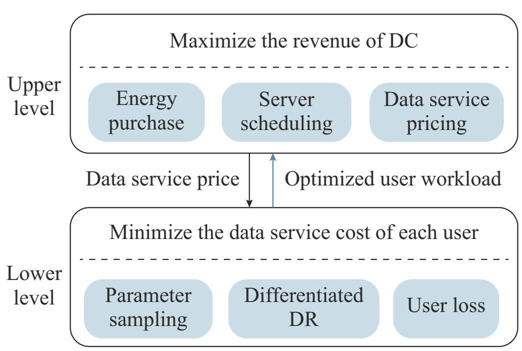

Fig. 4. Scheme of proposed optimal pricing problem of DC modeled as a Stackelberg game.

In the subsequent solving process, the upper-level problem is transformed into a mixed-integer linear programming (MILP) problem, and the lower-level problem is transformed into an integer linear programming (ILP) problem. Thus, it can be simply proven that an equilibrium point exists in the proposed Stackelberg game using the method described in [24].

A. Upper-level Problem

In this subsection, the DC optimization problem is formulated. Given the energy prices and user's DR characteristics, the DC sets data service prices with the goal of maximizing profits to optimize the demand for data services during each time period.

1) Objective Function

The objective of the DC is to maximize its profit, which consists of two components, as described by:

$$\max u_1 = \underbrace{\sum_{t=1}^{T} C_t^{\text{data}} D_t \Delta t}_{\text{Revenue of providing data services}} - \underbrace{\left[\underbrace{\sum_{t=1}^{T} C_t^{\text{e}} P_t^{\text{e}} \Delta t}_{\text{Electricity cost of grid}} + \underbrace{\sum_{t=1}^{T} C^{\text{RES}} P_t^{\text{RES}} \Delta t}_{\text{Cost of RES}} + \underbrace{C^{\text{GC}}\left(W^0 - W^{\text{RES}}\right)}_{\text{Cost of GC}}\right]}_{\text{Energy purchase cost}} \tag{15}$$

where u_1 is the objective function of the DC; Δt is the time interval; C_t^{e} is the electricity price at time t; P_t^{e} is the power of electricity that the DC purchases from the grid at time t; C^{RES} is the price of the RES; P_t^{RES} is the amount of RES power consumed by the DC at time t; C^{GC} is the price of the green certificate (GC); W^0 is the required consumption of RES generation over a period of time; and W^{RES} is the actual consumption of RES generation during a time period.

2) Constraints for Energy Purchase

In this paper, it is assumed that the DC, as a large local energy consumer, is given priority in purchasing electricity from a renewable generation company (RGC). In addition, the DC purchases electricity from the grid only when the power supply of the RGC is insufficient. Furthermore, the DC is committed to collecting a certain number of generation companies (GCs) and must purchase additional GCs when the accommodation of RES generation is insufficient.

$$P_t^{\text{DC}} = P_t^{\text{RES}} + P_t^{\text{e}} \tag{16}$$

$$W^{\text{RES}} = \sum_{t=1}^{T} P_t^{\text{RES}} \tag{17}$$

$$0 \le P_t^{\text{RES}} \le P_t^{\text{RES,max}} \tag{18}$$

$$0 \le P_t^{\text{e}} \le P_t^{\text{DC}} \tag{19}$$

where $P_t^{\text{RES,max}}$ is the maximum power output of the RES at time t.

3) Constraints of Server Scheduling

The constraints on server scheduling within the DC are given in (2)-(9), including the calculation of DC energy consumption, SLA, and operating constraints of the server.

4) Constraints of Data Service Price

The data service price mechanism is constructed using the DC through the following constraints. Equations (20)-(24) describe how the DC service pricing C_t^{data} is represented as a step function at time t. This step function contains N^{c} steps (levels), whose values are $S_1^{\text{data}} > S_2^{\text{data}} > ... > S_{N^{\text{c}}}^{\text{data}}$. Equation (23) ensures that the difference between adjacent price levels is sufficiently large for price discrimination.

Equation (24) imposes a constraint such that the average price of data services should be equal to the fixed price C^{fixed} before applying the new pricing scheme. This ensures that the variability of the proposed nonlinear pricing scheme is limited.

$$C_t^{\text{data}} = \sum_{i \in N^{\text{c}}} X_{t,i}^{\text{data}} S_i^{\text{data}} \tag{20}$$

$$\sum_{i \in N^{\text{c}}} X_{t,i}^{\text{data}} = 1 \tag{21}$$

$$C^{\text{data,min}} \le S_i^{\text{data}} \le C^{\text{data,max}} \tag{22}$$

$$S_i^{\text{data}} - S_{i+1}^{\text{data}} \ge \Delta C \quad i \in [1, N^{\text{c}} - 1] \tag{23}$$

$$\frac{1}{T}\sum_{t=1}^{T} C_t^{\text{data}} = C^{\text{fixed}} \tag{24}$$

where $X_{t,i}^{\text{data}}$ is a binary variable to designate the specific level i at time t from the step function; $C^{\text{data,max}}$ and $C^{\text{data,min}}$ are the maximum and minimum data service prices, respectively; and ΔC is the minimum difference between adjacent price levels.

B. Lower-level Problem

In this subsection, the DR model of DC users is established to respond to the time-varying pricing scheme of the DC.

1) Objective Function

The objective is to minimize users' total cost of the data service with the given data service price.

$$\min u_2 = \sum_{j=1}^{N^{\text{user}}} C_j^{\text{deal}} D^{\text{fix}} \tag{25}$$

where u_2 is the objective function of the users; and D^{fix} is the data service demand of each user (it is assumed that various users have the same load demand of the data service).

2) Constraints

The constraints on user DR in the DC are given in (10)-(14), which mainly include the user's WTP, delay tolerance, and the condition of lost users.

3) Transformation of Problem

To convert the bi-level optimization model into a single-level one, the lower-level problem, which includes (10)-(14) and (25), is rewritten into (26)-(31) to eliminate the non-continuous 0/1 variables in the model.

$$\min \tilde{u}_2 = \underbrace{\sum_{j=1}^{N^{\text{avail}}}\left[\sum_{t=1}^{T}\left(C_t^{\text{data}} + \varepsilon_{j,t}\right) D_{j,t}^{\text{avail}} \Delta t + \left(C_j^{\max} + \varepsilon^{\max}\right) D_j^{\text{loss}}\right]}_{\text{Price-sensitive users that can be scheduled}} + \underbrace{\sum_{t=1}^{T} C_t^{\text{data}} D_t^{\text{base}} \Delta t}_{\text{Price-insensitive users}} \tag{26}$$

where N^{avail} is the total number of price-sensitive users that can be rescheduled; $\varepsilon_{j,t}$ is the user preference of price-sensitive user j at time t; $D_{j,t}^{\text{avail}}$ is the data service demand of price-sensitive user j at time t; $\varepsilon^{\max}$ is the maximum preference of the user; D_j^{loss} is the data service demand of user j that is lost by the DC; and D_t^{base} is the total demand of price-insensitive users at time t, which is the basic workload.

The relevant constraints are expressed as:

$$D_t = \sum_{j=1}^{N^{\text{avail}}} D_{j,t}^{\text{avail}} + D_t^{\text{base}} \tag{27}$$

$$D_{j,t}^{\text{avail}} = T_{j,t}^{\text{avail}} D_{j,t}^{\text{a}} \tag{28}$$

$$D_j^{\text{loss}} + \sum_{t=1}^{T} D_{j,t}^{\text{avail}} = D^{\text{fix}} \tag{29}$$

$$0 \le D_{j,t}^{\text{avail}} \le D^{\text{fix}} \tag{30}$$

$$0 \le D_j^{\text{loss}} \le D^{\text{fix}} \tag{31}$$

where $T_{j,t}^{\text{avail}}$ is the status variable indicating whether user j is within the adjustable time range at time t; and $D_{j,t}^{\text{a}}$ is an intermediate variable of price-sensitive user j at time t.

Equations (26) and (29)-(31) ensure that when the data service price is given, users will choose the time period with the lowest price while meeting the time requirement. Equation (28) ensures that users choose data service prices only in an adjustable range of time.

Note that $\varepsilon_{j,t}$ is designed to help users make choices when the price is the same during several time periods, and $\varepsilon^{\max}$ is used to ensure that users choose to accept the price of data services at time t instead of user losses.

C. Solution Procedure of Bi-level Problem

After the lower-level problem is converted to a continuous optimization one, its KKT optimality conditions can then be obtained to transform the original bi-level problem into a single-level one. However, it is difficult to solve the single-level problem because of the nonlinearity in (15), where C_t^{data} and D_t are both decision variables. As shown in (32) and (33), the product terms can be replaced by linear components based on the KKT optimality conditions obtained in the lower-level problem using the strong duality theory [25].

$$\sum_{t=1}^{T} C_t^{\text{data}} D_t \Delta t = \sum_{j=1}^{N^{\text{avail}}}\sum_{t=1}^{T} C_t^{\text{data}} D_{j,t}^{\text{avail}} \Delta t + \sum_{t=1}^{T} C_t^{\text{data}} D_t^{\text{base}} \Delta t \tag{32}$$

$$\sum_{j=1}^{N^{\text{avail}}}\sum_{t=1}^{T} C_t^{\text{data}} D_{j,t}^{\text{avail}} \Delta t = -\sum_{j=1}^{N^{\text{avail}}} \lambda_j^{1,2} D^{\text{fix}} - \sum_{j=1}^{N^{\text{avail}}}\sum_{t=1}^{T} \mu_{j,t}^{1,2} D^{\text{fix}} - \sum_{j=1}^{N^{\text{avail}}} \mu_j^{1,4} D^{\text{fix}} - \sum_{j=1}^{N^{\text{avail}}}\sum_{t=1}^{T} \varepsilon_{j,t} D_{j,t}^{\text{avail}} - \sum_{j=1}^{N^{\text{avail}}}\left(C_j^{\max} + \varepsilon^{\max}\right) D_j^{\text{loss}} \tag{33}$$

where $\lambda_j^{1,2}$ is the dual variable of (29); and $\mu_{j,t}^{1,2}$ and $\mu_j^{1,4}$ are the dual variables of (30) and (31), respectively.

Thus, the objective function of the single-level problem can be rewritten as:

$$\max \tilde{u}_1 = -\sum_{j=1}^{N^{\text{avail}}} \lambda_j^{1,2} D^{\text{fix}} - \sum_{j=1}^{N^{\text{avail}}}\sum_{t=1}^{T} \mu_{j,t}^{1,2} D^{\text{fix}} - \sum_{j=1}^{N^{\text{avail}}} \mu_j^{1,4} D^{\text{fix}} - \sum_{j=1}^{N^{\text{avail}}}\sum_{t=1}^{T} \varepsilon_{j,t} D_{j,t}^{\text{avail}} - \sum_{j=1}^{N^{\text{avail}}}\left(C_j^{\max} + \varepsilon^{\max}\right) D_j^{\text{loss}} + \sum_{t=1}^{T} C_t^{\text{data}} D_t^{\text{base}} \Delta t - \sum_{t=1}^{T} C_t^{\text{e}} P_t^{\text{e}} \Delta t - \sum_{t=1}^{T} C^{\text{RES}} P_t^{\text{RES}} \Delta t - C^{\text{GC}}\left(W^0 - W^{\text{RES}}\right) \tag{34}$$

Equation (34) is subject to constraints (2)-(9), (16)-(24), and the linearized KKT conditions (26)-(31) of the lower-level problem. Finally, the preceding nonlinear problem is transformed into an MILP problem.

The GUROBI solver in MATLAB is employed to solve the MILP problem, similar to the method employed in [26]. The decision variables are summarized as follows.

1) Upper-level problem: C_t^{data}, C_i^{data}, $K_{n,t}^{\text{server}}$, n_t, P_t^{DC}, P_t^{e}, P_t^{RES}, $SX_{n,t}$, W^{RES}, $X_{n,t}^{\text{server}}$, and $X_{t,i}^{\text{data}}$.

2) Lower-level problem: D_t, $D_{j,t}^{\text{avail}}$, D_j^{loss}, and $D_{j,t}^{\text{a}}$.

IV. Case Studies and Simulation Results

A. Case Description

To validate the effectiveness of the proposed approach, a case study is employed that includes a DC and its users in the USA. An hourly interval is employed in the day-ahead optimization, and it is assumed that data service demand arrives hourly. The DC processes 0.9 million requests every minute during peak periods, as assumed in [27]. As shown in Fig. 5, the real-time pricing (RTP) data of Illinois is used as the predicted day-ahead electricity price. As shown in Fig. 6, a local RGC is set to provide a DC of up to 19 MW of photovoltaic (PV) output. The normalized original workload of the DC is shown in Fig. 7, which is based on a real-world Google workload trace [28]. Detailed parameters of the DC and its users are listed in Table I [22], [27], [29] and Table II [27], [28], respectively, and the other parameters are listed in Table III.

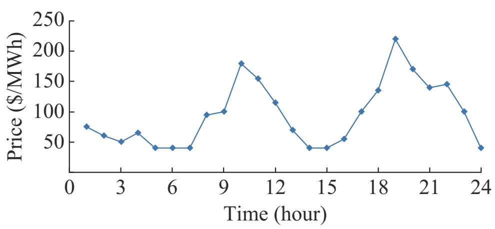

Fig. 5. Predicted electricity price in day-ahead market.

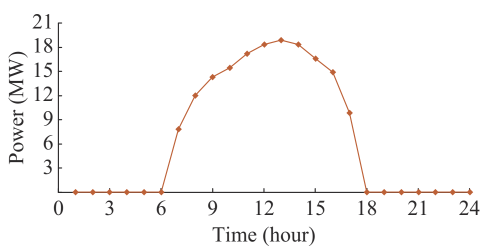

Fig. 6. Predicted values of PV output.

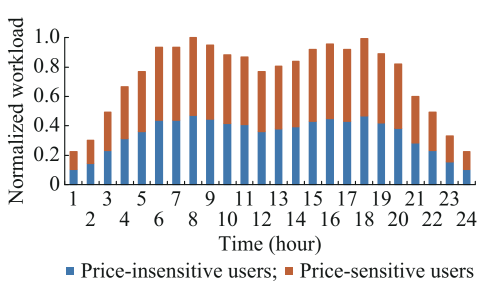

Fig. 7. Normalized original workload of DC.

It should be noted that the DC does not require all data of users, but mainly collects two key sets of user data, including the user data service period and WTP. The method of data collection can include user questionnaires or a limited collection of historical data after obtaining users' authorization. In addition, the DC is obliged to keep the collected data confidential and to prevent leakage. In the process of data analysis, data desensitization is required, i.e., some sensitive information is deformed through desensitization rules to achieve the reliable protection of sensitive and private data.

TABLE I
DC Parameters

Symbol	Parameter	Value
P^{peak}	Peak power of a server (W)	750
P^{idle}	Idle power of a server (W)	400
μ	Service rate of a server	4
η	PUE of DC	1.75
$d^{\max}$	The maximum delay time (s)	0.35
$n^{\max}$	The maximum number of servers	20000
T^{d}	The shortest interval between start and stop of servers (hour)	2
N^{c}	Number of data service price levels	3
C^{fixed}	Fixed data service price ($)	6×10^{-5}
$C^{\text{data,max}}$	The maximum data service price ($)	8×10^{-5}
$C^{\text{data,min}}$	The minimum data service price ($)	4×10^{-5}
ΔC	The minimum difference between adjacent price levels ($)	1×10^{-6}

TABLE II
User Parameters

Symbol	Parameter	Value
$C_j^{\max}$	The maximum of WTP of user j ($)	$N(7.5, 2^2)\times10^{-6}$
$t_j^{\max}$	The maximum of time adjustment of user j (hour)	$N(1.5, 2^2)$
D^{fix}	Data service demand of each user	1
	Percentage of price-sensitive users (%)	53.0
	Percentage of price-insensitive users (%)	47.0
	Percentage of time-sensitive users among price-sensitive users (%)	25.3
	Percentage of time-insensitive users among price-sensitive users (%)	74.7

TABLE III
Other Parameters

Symbol	Parameter	Value
C^{RES}	Price of RES ($/MWh)	40
C^{GC}	Price of a GC ($)	20
W^0	Required consumption of RES generation (MWh/day)	150

The following four scenarios are considered for comparison.

1) Scenario 1: the DC provides a fixed price for data services without RES accommodation.

2) Scenario 2: the DC provides a fixed price for data services while giving priority to RES accommodation.

3) Scenario 3: the DC provides a time-varying price for data services without RES accommodation.

4) Scenario 4: the DC provides a time-varying price for data services while giving priority to RES accommodation.

B. Simulation Results

The time-varying data service price and energy consumption optimization results of the DC are shown in Figs. 8-12 and Tables IV and V, in which the effects of data service price are analyzed in detail. Comparisons between scenarios are presented as below.

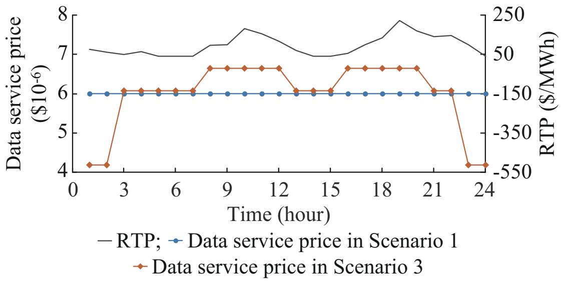

Fig. 8. RTP and data service prices in Scenarios 1 and 3.

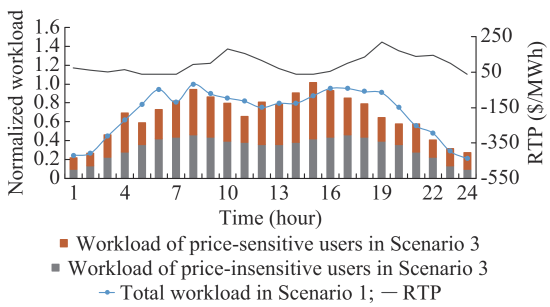

Fig. 9. RTP and workload in Scenarios 1 and 3.

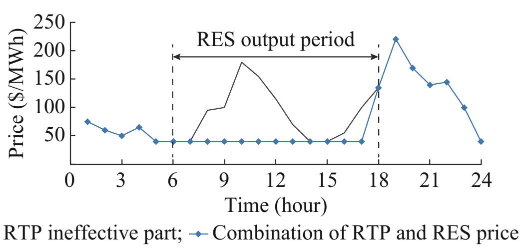

Fig. 10. Equivalent energy price curve.

1) Transmission from RTP to Data Service Price

Comparisons between Scenarios 1 and 3 are shown in Figs. 8 and 9, where the transmission effects of data service prices on RTP are analyzed.

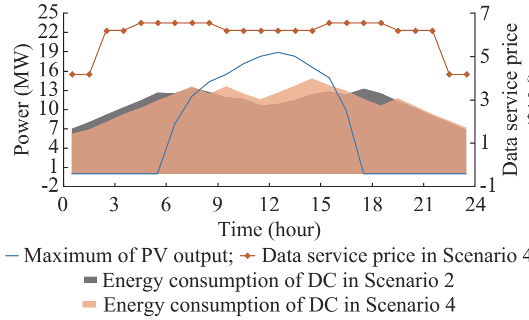

Fig. 11. Comparisons of RES accommodation in Scenarios 2 and 4.

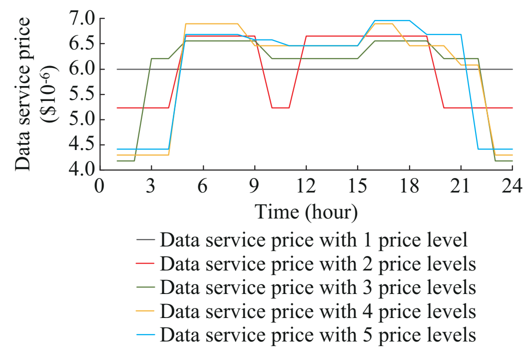

Fig. 12. Optimization results for time-varying data service price with different numbers of price levels.

As shown in Fig. 8, the peak and low value periods of the data service price in Scenario 3 are basically consistent with RTP, which is the result of the DC weighing user DR and user loss. The price-sensitive users in Scenario 3 change the time of the data service demand from the peak-price to the low-price period of the RTP to minimize the cost, as shown in Fig. 9. This also verifies the effectiveness of the DR model.

It can be concluded from the optimization results that the proposed time-varying data service price can effectively transfer the signals of variable energy prices to the users by setting a reasonable data service price. In addition, the optimized data service price can enable price-sensitive users to change their data service time and realize the DR potential of the users effectively.

2) Impact on RES Accommodation

Figures 10 and 11 show the effects of the proposed time-varying data service price mechanism on the RES accommodation when comparing Scenarios 2 and 4.

In Scenarios 2 and 4, the DC considers the purchase of RES and electricity from the grid when the RES falls short.

TABLE IV
REVENUE, COST DETAILS, AND OTHER ECONOMICS OF DC OPERATION

Scenario	Total revenue ($/day)	Total cost ($/day)	Profit ($/day)	Electricity purchase cost from the grid ($/day)	RES cost ($/day)	GC cost ($/day)	RES accommodation ratio (%)	User loss (%)
1	56751.8	26793.1	29958.7	26793.1				0
2	56751.8	19436.7	37315.1	13985.8	4901.8	549.1	75.2	0
3	56585.0	23788.1	32796.9	23788.1				8.88
4	57171.2	18963.5	38207.7	13263.1	5400.8	299.6	82.8	5.13

TABLE V
OPTIMIZATION RESULTS WITH DIFFERENT NUMBERS OF PRICE LEVELS

Number of price levels	Total revenue ($/day)	Electricity purchase cost from grid ($/day)	RES accommodation ratio (%)	User loss (%)
1	56752	13986	75.2	0
2	55531	12315	80.1	7.63
3	57171	13263	82.8	5.13
4	56687	12619	84.1	7.62
5	56391	12510	85.1	8.88

Therefore, the optimized data service price in Scenario 4 is similar in part to the equivalent energy price shown in Fig. 10, which combines the RTP and the RES price. The low-price period of data services is set when the RES consumption is insufficient or the RTP is low. Considering the effects of RES and GC, the optimized data service price shown in Fig. 11 increases the length of the low-price period at noon compared with the data service price shown in Fig. 8. This change encourages users to use data services as much as possible during the peak period of RES output, which increases the RES accommodation ratio from 75.2% to 82.8%.

With the expansion of the data service price range in the future, the proportion of price-sensitive users will continue to rise, which is currently at only 53.0%. This will further enhance the DC capability for RES recommendation.

3) Analysis of Economics

The revenue, cost details, and other economics of DC operation as reflected in Scenarios 1-4 are analyzed and presented in Table IV. It is noted that the DC maintenance and management costs and other fixed expenses are not calculated in this paper, as they are hardly affected by the price mechanism.

A comparison of Scenarios 1 and 3 reveals that the proposed time-varying data service price mechanism has a relatively small effect on the total revenue. However, it has significantly reduced the energy costs by as much as 11.2%, thus increasing the DC's profit. The main reason for the change in profit is that the DC always achieves a balance between increasing the price of data services and avoiding user losses. The reduction in total cost is due to the fact that more price-sensitive users choose to use data services during the low-price period of RTP after optimization.

Regarding the scenarios with RES, as can be observed from the results of Scenarios 2 and 4, the costs of purchasing electricity from the grid and GC are drastically reduced by 5.2% and 45.4%, respectively, which contribute to a reduction in the total cost and an increase in RES accommodation.

C. Sensitivity Analysis of Price Mechanism

A sensitivity analysis for the number of price levels is presented in this subsection, which greatly affects the ability of a DC to schedule users. The maximum number of data service price levels is set to be 5, as more price levels will result in difficulties in DC decision-making and increase the risk derived from the forecasting errors of DR. The optimization results are presented in Fig. 12 and Table V.

As shown in Fig. 12, with an increase in the number of data service prices, the similarity between data service prices and RTP continues to rise, which allows DCs to have a more flexible means of scheduling users with a WTP. In addition, the RES accommodation ratio as shown in Table V reveals that more price-sensitive users are scheduled to use data services during the peak period of RES output under the time-varying data service price mechanism with more price levels.

When the optimization results in Fig. 12 and Table V are combined, it can be concluded that the proposed time-varying data service price plays a positive role in improving the economic performance and environmental sustainability of the DC operation. Specifically, when the number of price levels increases from 3 to 5, the cost of purchasing electricity from the grid continues to decrease, while the RES accommodation ratio continues to increase. It is shown that when a DC has a greater number of price levels, it can formulate a more reasonable data service price mechanism based on the distribution of user's WTP to ensure the benefits of the DC and the RES accommodation ratio.

It should be noted that in our case studies, the total revenue and electricity purchase cost from the grid decrease simultaneously when the number of price levels increases from 1 to 2. This can be explained as follows. When the data service price involves only two price levels, the DC has to use a higher price to drive users' demands to the peak period of RES output where the lower data service prices are set. This leads to 7.63% user loss and 4.9% more RES accommodation ratio. In addition, as shown in Table V, when the number of price levels continues to increase (e.g., from 3 to 5), the price range of data services inevitably increases, which leads to a small increase in user losses and a small decrease in the total revenue of DC.

V. CONCLUSION

In this paper, a time-varying pricing scheme for a DC data service considering DR and RES accommodation is presented. Under the proposed price mechanism, the user price and time sensitivities can be analyzed and used to design a DR program, which help the DC schedule its workload more flexibly in managing its energy costs. A Stackelberg game between the DC and the users is also developed in this paper, where the upper- and lower-level problems aim to maximize the profit of the DC and to minimize the user data service cost, respectively. The bi-level optimization problem is transformed into a single-level MILP problem using KKT optimality conditions and the strong duality theory to solve the problem more efficiently. Case studies are conducted in four scenarios in a daily time range. Simulation results show that the proposed time-varying data service price mechanism plays a vital role in the economics of DC operation and in RES accommodation. DCs should set price levels for appropriate quantities to balance the economic and environmental benefits as well as user losses more effectively.

In fact, this paper shows that the users' WTP and time sensitivity cannot be decoupled completely, which means that a user's adjustable time range may also change with the prices. Future works will focus on the coupling relationship between user WTP and time sensitivity to ensure that the price mechanism is more reasonable and effective. In addition, the uncertainty of the RES output and DR will be analyzed.

REFERENCES

[1] H. Niu, L. Yang, J. Zhao *et al.*, "Flexible-regulation resources planning for distribution networks with a high penetration of renewable energy," *IET Generation, Transmission & Distribution*, vol. 12, no. 18, pp. 4099-4107, Dec. 2018.

[2] S. He, H. Gao, H. Tian *et al.*, "A two-stage robust optimal allocation model of distributed generation considering capacity curve and real-time price based demand response," *Journal of Modern Power Systems and Clean Eenrgy*, vol. 9. no. 1, pp. 114-127, Jan. 2021.

[3] Y. Shen, Y. Li, Q. Zhang *et al.*, "Consumer psychology based optimal portfolio design for demand response aggregators," *Journal of Modern Power Systems and Clean Eenrgy*, vol. 9. no. 2, pp. 431-439, Jan. 2021.

[4] L. Ni, W. Liu, F. Wen *et al.*, "Optimal operation of electricity, natural gas and heat systems considering integrated demand responses and diversified storage devices," *Journal of Modern Power Systems and Clean Energy*, vol. 6, no. 3, pp. 423-437, May 2018.

[5] Y. Chen, X. Feng, Z. Li *et al.*, "Multi-stage coordinated operation of a multi-energy microgrid with residential demand response under diverse uncertainties," *Energy Conversion and Economics*, vol. 1, no. 1, pp. 20-33, Oct. 2020.

[6] L. Ni, F. Wen, W. Liu *et al.*, "Congestion management with demand response considering uncertainties of distributed generation outputs and market prices," *Journal of Modern Power Systems and Clean Energy*, vol. 5, no. 1, pp. 66-78, Jan. 2017.

[7] J. Liu, L. Zéphyr, and C. L. Anderson. "Optimal operation of microgrids with load-differentiated demand response and renewable resources," *Journal of Energy Engineering*, vol. 146, no. 4, p. 04020027, May 2020.

[8] A. Asadinejad and K. Tomsovic, "Optimal use of incentive and price based demand response to reduce costs and price volatility," *Electric Power Systems Research*, vol. 144, no. 2, pp. 215-223, Mar. 2017.

[9] H. Wang, J. Huang, X. Lin *et al.*, "Proactive demand response for data centers: a win-win solution," *IEEE Transactions on Smart Grid*, vol. 7, no. 3, pp. 1584-1596, May 2016.

[10] CIE. (2019, May). Green data center white paper. [Online]. Available: http://www.ictlce.com/?p=584

[11] CCID Academy for Industry and Information Technology. (2020, Jun.). Research report on the development of China's "new infrastructure". [Online]. Available: https://www.ccidgroup.com/info/1096/21459.htm

[12] Q. Sun, C. Wu, Z. Li *et al.*, "Colocation demand response: joint online mechanisms for individual utility and social welfare maximization," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 3978-3992, Dec. 2016.

[13] A. Beloglazov and R. Buyya. "Energy efficient allocation of virtual machines in cloud data centers," in *Proceedings of 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing*, Melbourne, Australia , May 2010, pp. 577-578.

[14] R. K. Sharma, C. E. Bash, C. D. Patel *et al.*, "Balance of power: dynamic thermal management for Internet data centers," *IEEE Internet Computing*, vol. 9, no. 1, pp. 42-49, Jan.-Feb. 2005.

[15] Z. Xiao, W. Song, and Q. Chen, "Dynamic resource allocation using virtual machines for cloud computing environment," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 6, pp. 1107-1117, Jun. 2013.

[16] J. Li, Z. Bao, and Z. Li. "Modeling demand response capability by internet data centers processing batch computing jobs," *IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 737-747, Mar. 2015.

[17] N. H. Tran, D. H. Tran, S. Ren *et al.*, "How geo-distributed data centers do demand response: a game-theoretic approach," *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 937-947, Mar. 2016.

[18] Z. Chen, L. Wu, and Z. Li, "Electric demand response management for distributed large-scale internet data centers," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 651-661, Mar. 2014.

[19] D. Paul, W. Zhong, and S. K. Bose, "Demand response in data centers through energy-efficient scheduling and simple incentivization," *IEEE Systems Journal*, vol. 11, no. 2, pp. 613-624, Jun. 2017.

[20] N. Liu, Z. Dong, and R. Rojas-Cessa, "Task scheduling and server provisioning for energy-efficient cloud-computing data centers," in *Proceedings of 2013 IEEE 33rd International Conference on Distributed Computing Systems Workshops*, Philadelphia, USA, Jul. 2013, pp. 226-231.

[21] W. Wei, F. Liu, and S. Mei, "Energy pricing and dispatch for smart grid retailers under demand response and market price uncertainty," *IEEE Transactions on Smart Grid*, vol. 6, no. 3, pp. 1364-1374, May 2015.

[22] Y. Guo, Y. Fang, and P. P. Khargonekar, "Optimal workload and energy storage management for cloud data centers," in *Proceedings of 2013 IEEE Military Communications Conference*, San Diego, USA, Nov. 2013, pp. 1850-1855.

[23] Z. Liu, M. Lin, A. Wierman *et al.*, "Greening geographical load balancing," in *Proceedings of IEEE/ACM Transactions on Networking*, San Jose, USA, Jun. 2011, pp. 233-244.

[24] N. Liu, M. Cheng, X. Yu *et al.*, "Energy-sharing provider for PV prosumer clusters: a hybrid approach using stochastic programming and Stackelberg game," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 8, pp. 6740-6750, Aug. 2018.

[25] M. Yazdani-Damavandi, N. Neyestani, M. Shafie-Khah *et al.*, "Strategic behavior of multi-energy players in electricity markets as aggregators of demand side resources using a bi-level approach," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 397-411, Jan. 2018.

[26] Q. Deng, C. Feng, F. Wen *et al.*, "Evaluation of accommodation capability for electric vehicles of a distribution system considering coordinated charging strategies," *Energies*, vol. 23, no. 16, p. 3056, Aug. 2017.

[27] H. Xu and B. Li, "Reducing electricity demand charge for data centers with partial execution," in *Proceedings of the 5th International Conference on Future Energy Systems*, New York, USA, Jun. 2014, pp. 51-61.

[28] Z. Sun, F. Kong, L. Xue *et al.*, "Intelligent joint spatio-temporal management of electric vehicle charging and data center power consumption," in *Proceedings of International Green Computing Conference*, Dallas, USA, Nov. 2014, pp. 1-8.

[29] M. Ghamkhari and H. Mohsenian-Rad, "Energy and performance management of green data centers: a profit maximization approach," *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 1017-1025, Jun. 2013.

Chenwei Jiang received the B.Eng. degree in electrical engineering and automation from Zhejiang University, Hangzhou, China, in 2019. He is currently pursuing the master's degree with the School of Electrical Engineering, Zhejiang University, Hangzhou, China. His research interests include demand response, multi-energy system optimization, and renewable energy integration.

Chung-Li Tseng received the Ph.D. degree in industrial engineering and operations research from the University of California-Berkeley, Berkeley, USA, in 1996. He is currently an Associate Professor of Operations Management at the Business School of The University of New South Wales (UNSW), Sydney, Australia. Prior to joining UNSW, he was on the faculty of the University of Maryland, College Park, USA, and the University of Missouri, Rolla, USA. He has edited or co-edited several special issues of peer-reviewed journals. His published work has appeared in Operations Research, The Energy Journal, Energy Economics, European Journal of Operational Research, Construction Management and Economics, and other journals. He is a member of The Institute for Operations Research and the Management Sciences (INFORMS) and The American Society of Civil Engineers (ASCE), and a senior member of IEEE. He is a past President of the Section on Energy, Natural Resources, and the Environment (ENRE) of the INFORMS. He is the Editor-in-Chief of the ASCE Journal of Energy Engineering, and has served on the Editorial Board of ASCE Journal of Infrastructure Systems, International Journal of Electronic Business Management, and International Journal of Business Analytics. His research interests include operation management, sustainability, financial engineering, and project management.

Yizheng Wang received the B. E. degree in from Chongqing University, Chongqing, China, in 2018, and the M.E. degree from Zhejiang University, Hangzhou, China, in 2021, both in electrical engineering. He is currently working in the Economic and Technology Research Institute, State Grid

Zhejiang Electric Power Co., Ltd., Hangzhou, China. His research interests include multi-energy system optimization and energy management.

Zhou Lan is a Senior Engineer in the Economic and Technology Research Institute, State Grid Zhejiang Electric Power Co., Ltd., Hangzhou, China. His research areas include power system planning, power system stability analysis, and characteristic analysis of renewable energy generation.

Fushuan Wen received the B.E. and M.E. degrees from Tianjin University, Tianjin, China, in 1985 and 1988, respectively, and the Ph.D. degree from Zhejiang University, Hangzhou, China, in 1991, all in electrical engineering. He joined the faculty of Zhejiang University in 1991, and has been a Full Professor in electrical engineering since 1997. He is also a part-time Distinguished Professor in electrical engineering under Yusheng Xue Education Foundation in Hangzhou Dianzi University, Hangzhou, China. He had been a University Distinguished Professor in electrical engineering, the Deputy Dean of the School of Electrical Engineering, and the Director of the Institute of Power Economics and Electricity Markets in South China University of Technology, Guangzhou, China, from 2005 to 2009. He is a Professor in electrical engineering with the Department of Electrical Power Engineering and Mechatronics, Tallinn University of Technology, Tallinn, Estonia. He is the Editor-in-Chief of Energy Conversion and Economics (IET, Wiley), and a Deputy Editor-in-Chief of Automation of Electric Power Systems. He also serves as the Editor, Subject Editor, and Associate Editor of a few international journals. His research interests include power industry restructuring, power system alarm processing, fault diagnosis and restoration strategies, as well as smart grids and electric vehicles.

Fei Chen is a Professorate Senior Engineer in the Economic and Technology Research Institute, State Grid Zhejiang Electric Power Co., Ltd., Hangzhou, China. His research areas include power grid planning and design.

Liang Liang is a Senior Engineer in the State Grid Zhejiang Electric Power Corporation Jiaxing Power Supply Company, Jiaxing, China. His research areas include power system planning and construction.