# Data center demand response potential assessment considering multiple types of flexible resources

Hui Xu, Yong Li, Hui Zhu, Ji Wang, Chenghao Hou*
State Grid Liaocheng Power Supply Company, State Grid Corporation of China, Liaocheng 252000, Shandong, China

## ABSTRACT

This paper presents a method toevaluate the potential of data centers to involvement in power demand response, which is crucial for maintaining the equilibrium of supply and demand in the grid, promoting clean energy consumption, and mitigating energy crises. The data center is a valuable demand response resource due to its flexible adjustment methods, large load volume, and strong controllability, facilitated by its spatio-temporal scheduling mechanism. The proposed method considers multiple types of flexible resources and includes an analysis of data center participation demand response mechanisms, an evaluation model based on time series simulation optimization operation, and practical examples to verify its effectiveness.

**Keywords:** Data center, flexible resources, demand response, potential assessment

## 1. INTRODUCTION

In recent years, the energy crisis has intensified, resulting in various regions of the world facing energy supply and demand contradictions, leading to a tight overall supply and demand of electricity[1]. Power demand response is a crucial tool in ensuring the equilibrium of supply and demand in the grid, promoting the consumption of clean energy, and alleviating the energy crisis. Therefore, it is essential to explore the latent capacity of user-side demand response resource adjustment to maintain a balance between power supply and demand[2]. Data centers can help evaluate the demand response potential, which can further improve the operation flexibility of the demand side of the power system. This can enrich the operation adjustment methods of the power grid, promote the development of the power market, and ensure the equilibrium of supply and demand in the grid.

Current research on data center demand response primarily focuses on optimizing dispatching strategies. One approach involves developing workload scheduling strategies aimed at minimizing power costs within the data center. Another approach is to utilize valley filling scheduling strategies. A third approach involves developing coordinated scheduling strategies between the data center and power grid with the goal of minimizing their combined operation cost. Currently, there is limited research on the demand response potential of data centers. The concept of demand response potential was introduced in Reference[3], but the study only focused on the impact of delay-tolerant data load scheduling on the data center's adjustment potential. In Reference[4], another study was conducted that examined the energy consumption of data centers and assessed their load regulation potential. However, this study only focused on the demand response potential derived from space scheduling for delay-sensitive data loads. In Reference[5], the traditional data center energy consumption model was reconstructed, and a demand response potential model was developed to describe the spatiotemporal transfer characteristics of data load. However, this model does not account for the demand response potential generated by auxiliary equipment in data centers.

This paper highlights the importance of data centers participating in power demand response, despite their high energy consumption. Data centers have immense potential in demand response, yet a systematic and complete evaluation system is lacking. Therefore, this study delves into the demand response potential of data centers, analyzes the mechanism of data center load adjustment participating in demand response, and establishes a model for evaluating data center demand response potential. The proposed evaluation model and method are verified by example.

*13869550933@139.com

Fourth International Conference on Computer Science and Communication Technology (ICCSCT 2023), edited by Cheng Siong Chin, Wenbing Zhao, Changbo Cheng, Proc. of SPIE Vol. 12918, 129181D · © 2023 SPIE · 0277-786X · doi: 10.1117/12.3009383

## 2. ANALYSIS OF DEMAND RESPONSE MECHANISM OF DATA CENTER

### 2.1 Data center architecture

Data centers consist of various equipment, including IT, refrigeration, and power distribution equipment. In practice, microgrids, consisting of renewable energy, energy storage systems, and traditional generators, are commonly used in data centers[6]. Figure 1 shows the fundamental architecture of a data center[7]. The adjustable workload and cooling system power offer significant demand response potential to data centers. Additionally, the microgrid's multiple power supply modes can also provide demand response potential. The data center contains several types of flexible resources, which can aid in achieving demand response, optimal scheduling, and energy management objectives[8].

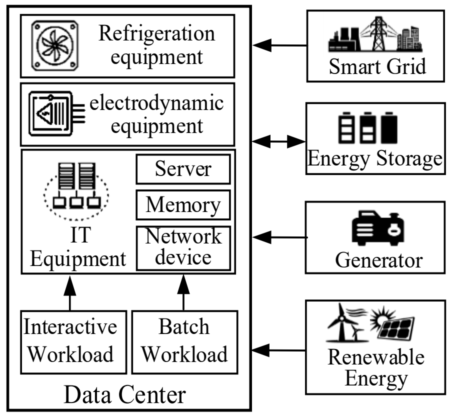


Figure 1. The fundamental architecture of a data center.

### 2.2 Data center demand response mechanism

The flexibility of demand response in data centers is demonstrated through its distinct mechanisms for scheduling workloads and regulating auxiliary equipment[9].

(1) Workload scheduling mechanism

Data center workloads are categorized into two types: interactive and batch workloads. Interactive workloads are space-scheduled and can be allocated by data center operators to any data center as long as load balancing is achieved. Batch processing workloads, on the other hand, are time-scheduled and do not require immediate processing upon arrival. These workloads have varying tolerable delay times depending on their delay tolerance levels and can be scheduled and processed at any time within the specified tolerance delay time. Therefore, the transfer of power load in data centers can be done on both spatial and time scales by transferring interactive workloads to other data centers and postponing the processing time of batch processing workloads, respectively.

(2) Mechanism of auxiliary equipment operation management

Data center auxiliary equipment comprises various systems such as refrigeration, lighting, new energy power generation, conventional generator set, and energy storage system. The refrigeration system plays a crucial role in maintaining the temperature inside a data center within normal range by utilizing the thermal inertia buffer of the building during low power consumption periods. This reduces the cooling load during peak power consumption periods, resulting in lower energy consumption for the data center. Additionally, conventional generator sets, energy storage equipment, and new energy power generation equipment can provide multiple sources of electric energy for the data center, enabling it to change its power access scheme and participate in demand response effectively[10].

## 3. DATA CENTER DEMAND RESPONSE POTENTIAL EVALUATION MODEL

The data center has abundant demand response resources on the user side, but the potential response is difficult to estimate. The application of various load adjustment mechanisms generates different demand response potentials with varying response time and duration. To assist data centers in determining their demand responsiveness in the electricity market while considering uncertain future tasks, this section models the demand response potential of data centers.

### 3.1 Data center model

The total energy consumption of the data center mainly consists of server equipment energy consumption, refrigeration system energy consumption, and other energy consumption, as in equation (1):

$$P_t = P_t^s + P_t^c + P_t^o \tag{1}$$

In equation (1): $P_t^s$ is the energy consumption of IT equipment; $P_t^c$ is the energy consumption of the refrigeration equipment; $P_t^o$ is the energy consumption of other devices.

The energy consumption of IT equipment is shown in equation (2):

$$P_t^s = P_{idle}^i m_t^i + \sum_{q=1}^{Q} P_{idle}^b m_t^{b,q} + \frac{P_{peak}^i - P_{idle}^i}{\mu^i} I_t + \frac{P_{peak}^b - P_{idle}^b}{\mu_t^b} \sum_{q=1}^{Q} B_t^q \tag{2}$$

In equation (2), $P_{idle}^i$ and $P_{peak}^i$ represent the idle power and peak power of a single server handling interactive workloads, respectively; $P_{idle}^b$ and $P_{peak}^b$ represent the idle power and peak power of a single server handling batch workloads, respectively; $m_t^i$ and $m_{t,q}^b$ are the number of active servers handling interactive workloads and batch workloads of type q, respectively; $I_t$ and $B_t^q$ represent the interactive workload to be processed and the batch workload of type q, respectively; $\mu_t^i$ and $\mu_t^b$ are the processing rates for interactive workloads and batch workloads, respectively.

The energy consumption of refrigeration equipment is shown in equation (3):

$$P_t^c = \frac{Q_h}{\eta_{cop}} + P_t^f \tag{3}$$

In equation (3): $Q_h$ is the thermal power of the data center; $\eta_{cop}$ is the performance coefficient of the refrigeration equipment; $P_t^f$ is the fan energy consumption of the refrigeration equipment.

### 3.2 Objective function

The calculation model for data center demand response potential aims to maximize the load response potential that a data center can generate during the demand response period. This is achieved by finding the difference between the initial load consumed by the data center from the grid and the load consumed by the data center from the grid through adjustment means. The objective function of the model is to maximize the demand response potential of the data center, which is shown in equation (4):

$$\max e_j^{\mathrm{DR}} = \max \left\{ \left| \sum_{t \in T} e_t^{grid} - \sum_{t \in T} e_t^{grid*} \right| \right\} \tag{4}$$

In equation (4): $e_j^{DR}$ is the demand response potential of the data center; $e_t^{grid}$ is the power supply of the power grid after the data center participates in demand response; $e_t^{grid*}$ is the power supply of the power grid when the data center does not participate in demand response.

### 3.3 Constraint condition

Load balancing conditions must be met when scheduling interactive workloads across multiple data centers. The constraint is represented by equation (5):

$$\sum_{n=1}^{N} I_{t,n} = \phi_t^I \tag{5}$$

In equation (5): $\phi_t^I$ is the interactive workload that reaches the data center.

Equation (6) shows the constraints that must be met when processing the batch workload within the specified delay time.

$$\sum_{t=1}^{\tau} B_t^q \le \sum_{t=1}^{\tau+D_q} \phi_t^{B,q} \tag{6}$$

In equation (6), $\phi_t^{B,q}$ is the amount of batch processing workload of class $q$ that arrives at the data center; $D^q$ is the maximum latency processing time for batch workloads of type $q$.

The service quality of a data center server can be adjusted to change the data load processing time. However, it is important to ensure that the service quality meets the requirement of processing the data load within a specified time. This constraint is represented by equation (7):

$$\frac{1}{m_t^i \mu^i - I_t} \le t^i \tag{7}$$

In equation (7), $t^i$ is the average processing latency of the workload.

Equations (8) and (9) impose data center server capacity constraints on servers that handle interactive and batch workloads.

$$0 \le m_t^i \le m_{\max}^i \tag{8}$$

$$\sum_{q=1}^{Q} \frac{m_{t,q}^b}{\mu_t^b} \le m_t^b \le m_{\max}^b \tag{9}$$

In equations (8) and (9): $m_{\max}^i$ is the maximum number of servers that can handle interactive workloads; $m_{\max}^b$ is the maximum number of servers that can handle batch workloads.

Equation (10) displays the constraints for the specified temperature range being maintained within the data center's internal environment.

$$T_{\min} \le T_{in} \le T_{\max} \tag{10}$$

In equation (10), $T_{in}$ is the indoor temperature of the data center; $T_{\max}$ is the maximum indoor temperature of the data center; $T_{\min}$ is the maximum indoor temperature of the data center.

Equation (11) represents the indoor temperature change process of the data center using an equivalent thermal parameter model based on the first law of thermodynamics.

$$\frac{\mathrm{d}T_{in}}{\mathrm{d}t} = \alpha(Q_h - Q_c) - \beta(T_{in} - T_{out}) \tag{11}$$

In equation (11), $\alpha$ and $\beta$ are the heat transfer coefficient of the internal building of the data center; $Q_h$ is the heat dissipation capacity of the data center; $Q_c$ is the cooling capacity of the data center.

In order to meet the demands of charging and discharging power and capacity, energy storage equipment must adhere to the limitations outlined in equation (12):

$$-P^{\text{bcmax}} \le P_t^b \le P^{\text{bdmax}} \tag{12}$$

In equation (12): $P^{\mathrm{bcmax}}$ is the maximum charging power of the energy storage equipment in the data center; $P^{\mathrm{bdmax}}$ is the maximum discharge power of the energy storage equipment in the data center.

Equations (13) and (14) are the data center capacity constraint adjustment:

$$L_t = L_{t-1} + P_t^b \eta^b \tag{13}$$

$$L^{\min} \le L_t \le L^{\max} \tag{14}$$

$L_t$ is the capacity of energy storage equipment in the data center; $L^{\max}$ and $L^{\min}$ are the upper and lower limits of energy storage capacity, respectively.

In order to ensure proper functioning, generators in data centers must adhere to both upper and lower output limits as well as climbing constraints, as outlined in equations (15) and (16):

$$P^{\mathrm{gmin}} \le P_t^g \le P^{\mathrm{gmax}} \tag{15}$$

$$R^{\min} \le P_t^{\mathrm{g}} - P_{t-1}^{\mathrm{g}} \le R^{\max} \tag{16}$$

In equations (15) and (16): $P^{\mathrm{gmax}}$ and $P^{\mathrm{gmin}}$ are the upper and lower limits of the data center generator, respectively; $R^{\max}$ and $R^{\min}$ are the upper and lower limits of the generator's climbing power, respectively.

# 4. EXPERIMENTAL SIMULATION AND RESULTS

## 4.1 Problem description

This paper focuses on a cloud computing data center operator that owns three large data centers distributed in different regions. The research specifically looks at data center 1 and its potential for demand response. Figure 2 shows the data center's task load arrival and renewable energy output. This paper aims to determine the demand response potential of a data center by analyzing three different scenarios. The first scenario involves calculating the theoretical demand response potential of the data center without considering any load regulation cost or electricity price. The second scenario takes into account the load adjustment cost and the time-of-use price signal to calculate the price-based demand response potential value. The third scenario considers the load adjustment cost and the incentive signal to calculate the potential for incentive-based demand response.

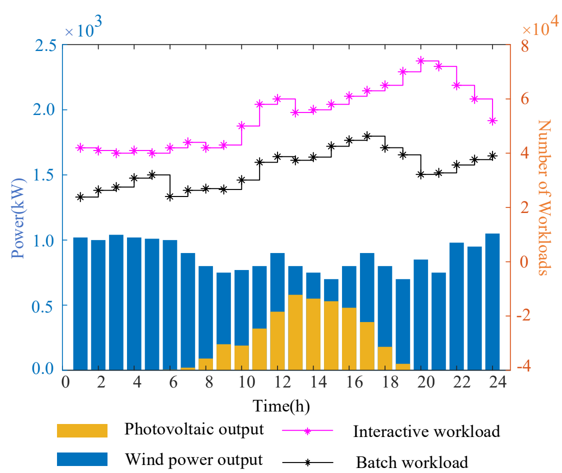


Figure 2. The data center's task load arrival and renewable energy output.

### 4.2 Simulation results and analysis

The evaluation results of data center demand response potential in each scenario are presented in Figure 3. The figure clearly shows that the data center has a significant potential for demand response, which can reduce the original load by about 50%, without considering the influence of economic factors. The space scheduling mechanism of the data center and the energy storage system did not participate in the demand response in each load regulation mechanism. This is because the delay-tolerant load only performs delayed processing on the time scale, which cannot reduce the total processing quantity, while the energy storage system has an initial and final capacity of 0, and the total power load cannot be reduced. The maximum demand response potential per hour is calculated, and it can be determined that the minimum load of the data center participating in demand response is 2.31 MW in each period, which is also the limit load for the data center to maintain the normal operation.

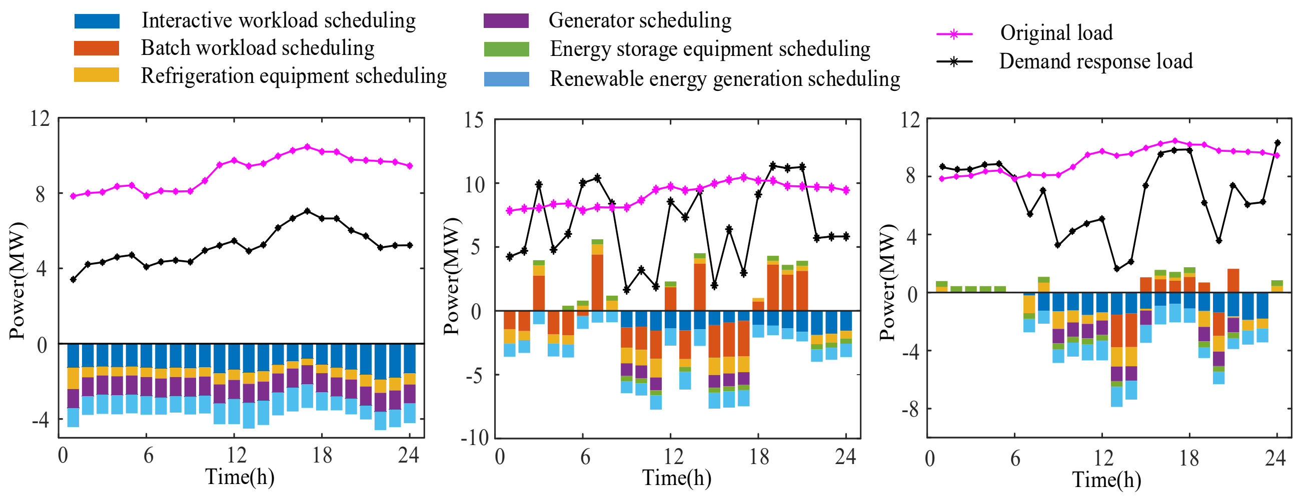


Figure 3. Data center demand response potential assessment results.

When participating in price-based demand response, data centers can optimize electricity costs by adjusting their load according to real-time electricity prices. During periods of low electricity prices, data centers can increase power load through three adjustment methods: time scheduling, refrigeration system adjustment, and energy storage system charging. This allows for more demand response potential during high electricity price periods. Conversely, during high electricity price periods, data centers can use six adjustment mechanisms, including time scheduling, control mechanism, space scheduling mechanism, and cooling system adjustment, to release their demand response potential. However, data centers will only adjust their load if the cost of load adjustment is lower than the cost of electricity price. Therefore, not all six adjustment mechanisms will be implemented in each period.

When the data center participates in the incentive demand response, it will adjust the load according to the adjustment direction of the incentive signal, optimize the electricity cost and obtain more demand response benefits. When the adjustment signal is up, the data center uses a time scheduling mechanism, refrigeration system adjustment, and energy storage system to increase electricity consumption. Although the electricity cost increases at this time, the demand response benefit obtained is greater than the electricity cost; when the adjustment signal is when it is down-regulated, the data center can reduce its own electricity cost through 6 kinds of adjustment mechanisms while increasing the down-regulation range of electricity consumption, which can not only reduce the electricity load and save electricity costs but also obtain demand response benefits; when there is no adjustment signal, although this situation does not require the data center to provide additional adjustment resources, the data center still uses the load adjustment mechanism to reduce electricity costs.

## 5. CONCLUSION

This paper analyzes the mechanism of data center involveding demand response with multiple load regulation mechanisms and the form of involveding in demand response and structures a demand response potential calculation model considering the operation characteristics of the data center, the characteristics of multiple load regulation mechanisms, and the form of participating in demand response. The example effectively calculates the potential value of the data center to participate in demand response under different scenarios. On this basis, the optimal dispatching strategy of each load regulation mechanism for the data center to participate in price-based demand response is obtained, and the maximum demand response potential is released under the premise of the lowest economic cost.

## REFERENCES


[1] Gao, C., Cao, X. and Yan, H., "Energy management of data center and prospect for participation in demand side resource scheduling," Automation of Electric Power Systems, 41(23), 1-7(2017).

[2] Yu, L., Jiang, T. and Zou, Y., "Price-sensitivity aware load balancing for geographically distributed internet data centers in smart grid environment," IEEE Transactions on Cloud Computing, 6(4), 1125-1135(2018).

[3] Li, J., Bao, Z. and Li, Z., "Modeling demand response capability by internet data centers processing batch computing jobs," IEEE Transactions on Smart Grid, 6(2), 737-747(2015).

[4] Chen, M., Gao, C. and Chen, S., "Bi-level economic dispatch modeling considering the load regulation potential of internet data centers," Proceedings of the CSEE, 39(05), 1301-1314(2019).

[5] Chen, M., Gao, C., Shahidehpour, M., et al., "Internet data center load modeling for demand response considering the coupling of multiple regulation methods," IEEE Transactions on Smart Grid, 23(3), 2060-2076(2021).

[6] Cupelli, L., Schutz, T., Jahangiri, P., et al., "Data center control strategy for participation in demand response programs," IEEE Transactions on Industrial Informatics, 14(11), 5087-5099(2018).

[7] Tran, N., Tran, D., Ren, S., et al. "How geo-distributed data centers do demand response: A game-theoretic approach," IEEE Transactions on Smart Grid, 7(2), 937-947(2016).

[8] Wang, H., Huang J., Lin, X., et al. "Proactive demand response for data centers: A win-win solution," IEEE Transactions on Smart Grid, 7(3), 1584-1596(2016).

[9] Dou, H., Qi, Y., Wei, W., et al. "Carbon-aware electricity cost minimization for sustainable data centers," IEEE Transactions on Sustainable Computing, 2(2), 211-223(2017).

[10] Ding, Z., Xie, L., Lu, Y., et al., "Emission-aware stochastic resource planning scheme for data center microgrid considering batch workload scheduling and risk management," IEEE Transactions on Industry Applications, 54(6), 599-608(2018).