A Two-stage Demand Response Strategy for Datacenters in the Smart Grid Environment

Yuling Li¹, Xiaoying Wang*, Peicong Luo¹, Dingchang Huang¹, Mengmeng Zhao¹

¹State Key Laboratory of Plateau Ecology and Agriculture, Department of Computer Technology and Applications, Qinghai University, Xining, China, 810016

*Correspondence: wxy_cta@qhu.edu.cn

Abstract. As a large load of the smart grid, datacenters usually have a huge demand for power. Datacenter participation in demand response programs of smart grids can help the grid to adjust the load while reducing its own power costs so as to save electricity costs. In this paper, we designed a two-stage demand response participation strategy to alleviate the peak load pressure of the grid and reduce the power cost of the datacenter. The first stage determines whether to participate in demand response according to the real-time electricity prices variation of the power grid and incentive information will be sent to encourage users to participate in the program to help shave the peak load. In the second stage, the datacenter interacts with its users by allowing users to submit bid information by auction. Then the datacenter selects the tasks of the winning users to postpone processing with awards, which reduces the electricity costs of the datacenter and effectively meets the demand response requirements of the smart grid. In this way, the goals of both parties could be achieved.

1. Introduction

In order to improve the reliability, stability, and sustainability of the smart grid, power companies in some countries have initiated a series of demand response (DR) programs to encourage their users to participate in with the aim of reducing the peak load of the grid and thus improving the reliability of the power grid [1]. In general, there are two main types of demand response programs of the smart grids, including incentive-based and price-based programs. In particular, incentive-based demand response programs include direct load control and interruptible load control. Price-based demand response programs include time-of-use electricity prices, real-time electricity prices, peak electricity prices and other price-related programs [2]. A great number of power companies are also trying to implement special demand response programs to achieve the goal of peak load shaving. Demand response programs also provide users with incentives to encourage user participation to adapt power consumption to demand response goals. In China, for example, Jiangsu province first experimented with a dynamic seasonal electricity pricing policy in 2015, using a potential way to encourage users to participate in demand response [3].

In this paper, we aim to reduce datacenter power costs by involving the datacenter to participate in the smart grid demand response program. We propose a two-stage demand response strategy. In the first stage, the power grid sends the DR signals to datacenter according to the variation of real-time electricity prices, while sending incentive information to the datacenter. The second stage is the bidding selection between the datacenter and its users. The user submits bidding information firstly, including the number of tasks that can be postponed and the cost that the datacenter needs to pay due
to some of the tasks are postponed execution. Then, the datacenter obtains the optimal solution by stochastic searching, delays the execution of the task of the winning users and rewards them. Using the two-stage strategy, the purpose of reducing the peak load pressure of the smart grid while reducing the power cost of the datacenter could be achieved.

2. Related Work
In recent years, the method of how datacenter participates in demand response programs is has been discussed a lot. Zhou et al. [4] used the real-time electricity price as a demand response signal in the deregulated electricity market and modeled datacenter decisions on power company selection and workload scheduling as a many-to-one game model. In [5], they adopted time-varying rewards to motivate users willing to participate in demand response projects. They used a game theory framework to model the game between a single datacenter and its users. Wang et al. [6] proposed a price-based incentive method. In order to motivate users to participate in their demand response projects, power companies set different power prices in different datacenters, encouraging users to shift the load to datacenters which with lower electricity prices. Wang et al. [7] regarded datacenter operations as a problem of minimizing energy costs. They designed a distributed demand response algorithm. Administrators send optimized messages by broadcasting messages to each data center, and then each datacenter itself optimizes. Guo et al. [8] aimed at the problem of a serious carbon footprint resulted from using a backup diesel generator to supply power when datacenter participating in demand response. They focus on the effective and environmentally-friendly demand response of the datacenter. Wang [9] et al designed a false data injection attack against nonlinear state estimation in a cyber-adversarial system and the system mainly focused on smart grid environment. Compared with these previous research works, the main contributions of this paper include:

• establishing interactions between smart grids, the datacenter, and users. Optimize them by designing a two-stage demand response strategy.
• using real-time electricity price variation as the demand response signal and designing a corresponding incentive mechanism.
• designing the second stage of the datacenter interaction mechanism with the users: users submit bid information to the datacenter voluntarily by the method of auction, and the datacenter selects some users by analyzing the optimal choice.

3. Problem Statement
In this section, we mainly introduce the interaction between the grid, the datacenter, and users. We aim at the problem of datacenter participating in demand response to establish the optimization problem. Assuming that the datacenter consists of $N$ hosts, denoted as host $1$ to host $N$. We also assume the electricity price at time $t$ is $\sigma(t) (\$/Wh)$, the energy consumption at time $t$ slot is $E(t)$ and the revenue brought by executing a task is $\gamma(S)$.

3.1. Smart grid consideration
The main purpose of this paper is to use the real-time electricity price of the smart grid, and send a demand response signal to the datacenter according to the electricity price level. Considering that the datacenter participates in demand response, the power grid should give the datacenter a certain reward based on the participation degree of $DC$. Here, we use the average value of the power consumption of the datacenter at each time slot in the past 5 days as the reference value $p_{base}$. If the actual power consumption reduction after the $DC$ participates in the demand response is $p_{red}$, then the execution rate is defined as $r = \frac{p_{red}}{p_{base}}$. Therefore, we refer to the reward method proposed in [10] and define the reward at time $t$ as Eq. (1), where the baseline is a constant, which is a base value for rewards.

$$
\theta(t) = \begin{cases} 
    r \cdot \text{baseline}, & 0 \leq r < 1 \\
    \text{baseline}, & r \geq 1 
\end{cases}
$$

(1)
3.2. Problems definition
In general, datacenter should guarantee the QoS (quality of service) of users when datacenter process the requests of users. In this section, we define a penalty model for the datacenter that cannot guarantee the QoS. If tasks submitted by users are postponed due to the datacenter participating in the demand response, the datacenter will assume punishment. We adopted the penalty model adopted in [11], using \( t_{sub} \) and \( t_{exec} \) represent the submission time and actual execution time of the task, respectively. The penalty for delaying task \( i \) can be defined as (2),

\[
\mu(i) = (t_{exec} - t_{sub}) \cdot \beta
\]

where \( \beta \) is a constant used to reflect the penalty.

In general, computing power consumption is the dominant part of the datacenter power consumption [12], and computing power consumption is mainly related to changes in load and frequency. Denote \( p_u \) as the computing power consumption of host \( n \) in the \( i^{th} \) slot. We use a linear calculation method to calculate power consumption [13], as shown in Eq. (3).

\[
P_u = p_{\text{max}} \cdot (c + (1-c) \cdot u_n)
\]

where \( p_{\text{max}} \) is the maximum power consumption of the server, \( c \) is the percentage of static power consumption, and \( u_n \) is the CPU utilization of the server \( n \). Hence, the total power consumption of datacenter at time \( t \) can be calculated as

\[
P_t = \sum_{n=1}^{N} P_u
\]

3.3. User-side management
In this paper, users voluntarily decide whether to participate in the demand response and we consider two type tasks, interactive type and batch type, respectively. Denote \( I_u(t) \) and \( B_u(t) \) as the number of interactive tasks and batch-type tasks respectively, which are submitted at time \( t \) slot by user \( u \).

If the user decides to participate in the demand response, the user submits the bid information in a two-tuple manner. The number of tasks \( \phi_u(t,r) \) that can be delayed to process from time \( t \) to \( r \) and the cost \( \alpha(u) \) resulted from postponing the execution of some tasks of user \( u \). We use the binary variable \( x_u \) to record whether the user \( u \) wins the bid. The value of \( x_u \) indicates that user \( u \) is selected or not.

3.4. Optimization problem definition
At first, we discuss the optimization problem of the first stage. The main goals of this stage are to improve the reliability of the power grid and reduce the peak load pressure. At this stage, the power grid will send demand response signals and incentive information to the datacenter based on the real-time electricity price.

In the second phase, we show the interaction between the datacenter and its users. We use the method of load shifting in a time dimension to reduce the electricity cost of datacenter. Denoting \( m \) is the number of users who submit tasks to the datacenter at time \( t \). The number \( \delta(t) \) of batch-type tasks executed at time \( t \) includes both the batch-type tasks submitted in time \( t \) and the batch-type tasks that were postponed from previous time slots.

\[
\delta(t) = \sum_{t' \leq t} I_u(t') - \sum_{t' \leq t} \phi_u(t,t') + \sum_{t' \leq t} \sum_{x \in \mathcal{X}} \phi_u(t,t') \cdot x
\]

where \( r' > t, t' > t \).

Denote \( \lambda(t) \) as the total amount of tasks that the datacenter needs to process at the time \( t \), which can be calculated by Eq. (6).

\[
\lambda(t) = \sum_{n=1}^{N} I_u(t) + \delta(t)
\]

Hence, denote \( \pi(t) \) and \( r_{\text{user}} \) as the number of tasks that violate QoS and the reward given to users by the datacenter, respectively. Therefore, \( r_{\text{user}} \) consists of the penalty \( \sum_{i=1}^{N} \mu(i) \) caused by the datacenter postponed the tasks to be processed of users in order to respond to the demand response, and the cost
\[ \sum_{u} \alpha(u) \cdot x_u \] paid to some users who actively bid. Therefore, the total cost of the datacenter at time slot \( t \) can be calculated as

\[
C(t) = E(t) \cdot \sigma(t) + r_w \cdot (\lambda(t) \cdot \gamma - \theta(t))
\] (7)

Overall, the goal of datacenter interaction with its users is to reduce total costs, which means the datacenter should participate in the demand response positively. The main goal of the second-stage optimization problem is to reduce the power consumption of the datacenter as much as possible during the \( DR \) response period, which means datacenter postpone some tasks to the time of low electricity price be processed based on the bid information of users. The objective of the optimization problem is to minimize the cost of the datacenter, which can be depicted as

Minimize \( C(t) \) (8)

4. Strategies and Methods

To address the issue mentioned in Section 3, we proposed a two-stage demand response strategy for cost reduction of datacenter and peak load shaving of the power grid and also implemented two other methods for comparison. Figure 1(a) and (b) show the real-time electricity price and historical 5-day average power consumption used in this paper, respectively. It can be clearly found that there are two peak electricity price periods, two time periods between 6:00-7:00 and 17:00-20:00 respectively.

Figure 1 Real-time electricity price and average power consumption

4.1 Two-stage demand response method (OP)

In this subsection, we will show in detail the issues involved in the proposed two-stage approach.

Stage I: the interaction between grid and datacenter

The main goals of this stage are to improve grid reliability and reduce peak load pressure. At this stage, the grid will send \( DR \) signals based on the price of electricity. In order to allow grid users to have sufficient preparation time to adjust their own loads to participate in demand response, we adopt the method of using real-time electricity prices variation to send \( DR \) signals. We also set an incentive method to send incentive information according to Eq. (1) so as to attract more users to participate in demand response.

Stage II: The datacenter interacts with its users

At this stage, we adopt reverse auction to manage tasks of users. Users submit bidding information to datacenter when the datacenter receives demand response signals. Then the datacenter selects the winning user, pays the required cost to them, and reallocates the tasks of these users according to the auction results.

Auction model: In this process, we regard the datacenter as the buyer and the user as the seller. Each user \( u \) voluntarily submits bid information in a two-tuple manner \( \langle \phi(u), \alpha(u) \rangle \). Then, the datacenter selects the winning user according to the objective function in Eq. (8). Finally, the datacenter reallocates the tasks of the winning user.
Solution: Since the optimization problem in the second stage is an NP-hard problem, it is similar to the knapsack problem. Therefore, we adopt the Genetic Algorithm (GA) to solve the optimization problem. The most important part of GA is the fitness function. In this paper, we define the fitness function as Eq. (8).

4.2 Best effort (BS)
In this strategy, the datacenter participates in the demand response plan of the grid at the first stage, but does not conduct user auctions and bidding processes at the second stage. At the second stage, the datacenter selectively delays some tasks only according to the needs of the datacenter, regardless of whether the tasks submitted by users can be postponed. Therefore, the datacenter may violate the QoS of the user and at the same time assume a certain penalty for delaying the execution of tasks of some users arbitrarily regardless of task characteristics.

4.3 Static method (ST)
In this strategy, the datacenter does not participate in the demand response programs and just manages the workload as usual.

5. Evaluation Results
In order to verify the effectiveness of the proposed method, this paper uses Cloudsim-plus tool to simulate the interaction between the datacenter and its users. In the experiment, we adopt the Google-trace workload for simulation, and we assigned tasks to 10 users.

Figure 2 shows the task scheduling conditions under the three strategies. Obviously, tasks under the ST strategy do not perform any scheduling. Compared with ST method, when using BS and OP, fewer tasks were processed during the six demand response time periods, such as at 6:00, 7:00, 17:00, 18:00, 19:00, and 20:00. This is because both BS and OP strategies are involved in demand response of stage I and task scheduling is also performed. However, the task reduction is much more under the OP strategy during some time slots.

We analyzed the total cost reduction situation of the BS and OP strategies relative to the ST strategy. Figure 3 details the relative reduction conditions under the three strategies, which is highest in the OP policy. Because the OP method participates in the demand response, and it postponed some tasks execution when the electricity price is high, thereby reducing the electricity cost included in the total cost. Compared with the BS strategy, since the OP strategy uses user bidding to participate in demand response and achieves a higher execution rate, the rewards obtained from the grid side are higher than BS strategy. Moreover, the BS strategy results in higher penalties because the datacenter randomly delays the tasks to be processed by some users.

Furthermore, we also compared the detailed rewards given to the datacenter by the grid at each DR moment under the BS and OP strategies and found that the BS strategy is lower than the OP strategy. In order to further prove the efficiency of the proposed method, we analyzed the power consumption
at each time slot under the three strategies. At the DR moments, the power consumption of the BS and OP strategies all are reduced, but the OP strategy was reduced even more. Overall, the demand response execution rate of the OP strategy is higher than that of the BS strategy. Therefore, the OP strategy can minimize the total cost of the datacenter.

6. Conclusions
In this paper, we proposed a two-stage datacenter demand response strategy. The first stage is the interaction between the smart grid and the datacenter, which mainly includes the grid sending demand response signals to the datacenter. In the second phase, the datacenter interacts with its users. Users submit bid information to the datacenter by the method of auction, the information includes the number of tasks can be postponed and the costs caused by postponing these tasks. Then, the datacenter delays the execution of tasks of the winning user and pays the relevant costs after the bid selection. Simulation experiments prove that the strategy proposed in this paper can well achieve the goal of participating in demand response, which can not only reduce the power cost of the datacenter but also reduce the peak load pressure of the power grid.

7. References
[1] Tran N H, Tran D H, Ren S How geo-distributed data centers do demand response: A game-theoretic approach. IEEE T SMART GRID, 2015, 7(2): 937-947.
[2] Albadi M H, El-Saadany E F. Demand Response in Electricity Markets: An Overview Power Engineering Society General Meeting, IEEE, 2007.
[3] Available on : https://max.book118.com/html/2018/0627/5201014120001300.shtm.
[4] Zhou, Zhi, F. Li, Z. Li Pricing Bilateral Electricity Trade between Smart Grids and Hybrid Green Datacenters. ACM, 2015, 43(1): 443-444.
[5] Zhan, Y., Xu, D., Yu, H., Yu, S Incentivizing users of data centers participate in the demand response programs via time-varying monetary rewards. arXiv preprint arXiv:1604.01950.
[6] Wang, H., Huang, J., Lin, X. Mohsenian-Rad, H, Proactive demand response for data centers: A win-win solution. IEEE T SMART GRID, 7(3), 1584-1596.
[7] Wang H, Ye Z Renewable energy-aware demand response for distributed data centers in smart grid, IGSEC. IEEE, 2016: 1-8.
[8] Chen, Hao, Michael C. Caramanis, and Ayse K Coskun Reducing the data center electricity costs through participation in smart grid programs, International Green Computing Conference. IEEE, 2014:1-10.
[9] Wang J, Hui L C K, Yiu H S M. System-state-free false data injection attack for nonlinear state estimation in smart grid[J]. International Journal of Smart Grid and Clean Energy, 2015, 169-176.
[10] Incentives method. Available on: https://www.taipower.com.tw/tc/index.aspx.
[11] Garg S K, Gopalaiyengar S K, Buyya R SLA-based resource provisioning for heterogeneous workloads in a virtualized cloud datacentre, International conference on Algorithms and architectures for parallel processing, 2011: 371-384.
[12] Liu y, Yang Errui, Xu Jianxia Energy Consumption Management in Cloud-Oriented Data Centers. Journal of Telecommunication Science, 2012(12):102-108. (In Chinese)
[13] Wang X, Du Z, Chen Y, et al A green-aware virtual machine migration strategy for sustainable datacenter powered by renewable energy. SIMUL MODEL PRACT TH, 2015, 58: 3-14.

Acknowledgments
This paper is supported by the National Natural Science Foundation of China (No. 61762074, No.91847302, No.61563044 and No. 61862053), and National Natural Science Foundation of Qinghai Province (No.2019-ZJ-7034, No.2017-ZJ-902).