

Electric Demand Response Management for Distributed Large-Scale Internet Data Centers

Zhi Chen, *Student Member, IEEE*, Lei Wu, *Member, IEEE*, and Zuyi Li, *Senior Member, IEEE*

Abstract—This paper evaluates the electric demand response (DR) management for distributed large-scale Internet data centers (IDCs) via the stochastic optimization approach. The electric DR of IDCs refers to the capability of optimally shifting cloud service tasks among distributed IDCs. Thus, the energy consumption reduction at certain IDC locations could be considered as the DR provision capacity in day-ahead DR programs. Cloud service tasks of IDCs include processing, storage, and computing tasks, which are further categorized into interruptible and non-interruptible tasks. The proposed model determines the optimal hourly DR capabilities of individual IDCs while considering uncertain coming cloud service tasks to individual IDCs. The major contribution of this paper is to rigorously formulate the DR capability of IDCs as changes in the electricity consumption when shifting cloud service tasks among distributed IDCs in different time zones, while considering the energy consumption for providing IT service, cooling, shifting cloud service tasks, environmental impacts, and uncertain coming tasks. The proposed model would enhance the financial situation and improve the environmental impacts of distributed IDCs by participating in day-ahead DR programs. The stochastic optimization adopts scenario-based approach via the Monte Carlo (MC) simulation for minimizing the total electricity cost, which is the expected electricity payment minus the revenue from the DR provision. The proposed model is formulated as a mixed-integer linear programming (MILP) problem and solved by state-of-the-art MILP solvers. Numerical results show the effectiveness of the proposed approach for solving the optimal electric DR management problem for distributed large-scale IDCs.

Index Terms—Distributed IDCs, environment, price-based demand response management, stochastic optimization.

Variables:

NOMENCLATURE

C^C	Cost of cooling (\$).
C^E	Cost of emission evaluated via carbon rates (\$).
C^{IT}	Cost of IT service (\$).
C^{SU}	Startup cost of a server cluster (\$).
C^T	Cost of transmitting tasks (\$).
d	Index of IDCs.
DR	DR capacity by shifting tasks (kW).
g	Index of discrete server frequency levels.

H_d^s	Processing hour requirement on servers in an IDC.
i	Index of IDC component clusters.
I	Commitment of a component cluster.
I^{DR}	Binary variable indicating if an IDC provides DR: 1 if it provides DR, 0 otherwise.
I^C	Commitment status of an air conditioner (AC): 1 if the AC is ON, 0 otherwise.
I^S	Binary variable indicating if a wavelength-division multiplexing (WDM) sends tasks out: 1 if it sends tasks out, 0 otherwise.
I^U	Binary variable indicating the temperature level: 1 if the temperature is above the maximum temperature level, 0 otherwise.
l	Index of optical cables.
p	Power consumption of an IDC component cluster (kWh).
p^C	AC power consumption (kW).
p^W	Power consumption of a WDM when sending out tasks (kW).
RE	Revenue of all IDCs from demand response (\$).
R^S	Number of per unit IT cloud service tasks shifted out of an IDC in each hour.
R^R	Number of per unit IT cloud service tasks shifted into an IDC in each hour.
s	Index of scenarios.
t	Index of time periods.
T^{sup}	Air conditioner supplied temperature (°F).
T^{out}	Indoor temperature (°F).
U	CPU utilization of an IDC server cluster.
y	Startup indicator of an IDC server cluster.
z	Shutdown indicator of an IDC server cluster.
λ	Number of cloud service per unit tasks arriving at an IDC.

Manuscript received December 20, 2012; revised March 14, 2013 and April 30, 2013; accepted June 05, 2013. Date of publication August 15, 2013; date of current version February 14, 2014. This work was supported in part by the U.S. National Science Foundation grant ECCS-1102064.

Z. Chen and L. Wu are with Electrical and Computer Engineering Department, Clarkson University, Potsdam, NY 13699 USA (e-mail: zhichen@clarkson.edu; lwu@clarkson.edu).

Z. Li is with Electrical and Computer Engineering Department, Illinois Institute of Technology, Chicago, IL 60616 USA (e-mail: lizu@iit.edu).

Digital Object Identifier 10.1109/TSG.2013.2267397

μ, γ	Binary variable indicating the status of sending/receiving tasks through optical cables.
Parameters:	
A, B	Power consumption coefficients of a server cluster (kW/GHz ³ , kW).
b	DR reward rate (\$/kWh).
c	Day-ahead electricity price announced by an Independent System Operator (ISO) in the day-ahead market (\$/kWh).
CM	Capacity margin of an IDC (tasks/h).
C_p	The specific heat of air (kJ/kg * K).
D	Response delay of a server cluster (h).
E_i	The energy saving of cluster i by shifting out one per-unit cloud service task (kWh per unit task).
f_d	Air flow rate between AC and component clusters of IDC d (m ³ /s).
FL_l^{\max}	Transmitting capacity of cable l (tasks/h).
M	Large positive number.
ND	Number of IDCs.
NG	Number of discrete server frequency levels.
NI_d	Number of computing clusters in IDC d .
NL_d	Number of cables connected to IDC d .
NS	Number of scenarios.
NT	Total number of hours in the schedule horizon.
P^{\max}	Maximum power level of an IDC component cluster (kW).
P^{\min}	Minimum power level of an IDC component cluster (kW).
PC^{\max}	Maximum power level of an AC cluster (kW).
PC^{\min}	Minimum power level of an AC cluster (kW).
PUE	Power utilization effectiveness of an IDC.
PT	WDM power consumption (kW).
q^{\max}	Maximum DR capability of an IDC (kW).
R_i	Original number of per unit IT cloud service tasks arriving at a local IDC for cluster i .
SU	Startup cost of server (\$).
T^{\max}	Maximum temperature level (°F).
T^{\min}	Minimum temperature level (°F).
U^{\max}	CPU utilization upper limit of server.
η_{ac}	Efficiency of an air conditioner.

ρ	Air density (kg/m ³).
β^s	Probability of scenario s .
ε	System inertia.
ϕ	Carbon emission rate (\$/kWh).
λ^{\max}	Maximum capacity of an IDC component cluster (tasks/h).
λ^{\min}	Minimum capacity of an IDC component cluster (tasks/h).
λ_g	Threshold of employing CPU frequency at level g (tasks/h).

Sets:

T_{IT}	Set of interruptible tasks.
T_{NIT}	Set of non-interruptible tasks.

I. INTRODUCTION

LARGE-SCALE internet data centers (IDCs) are increasingly emerging across the world for providing cloud services to meet the growing computing and storage requests. A typical data center has an electricity consumption level of tens of megawatts [1], which is equivalent to the electricity consumption of about ten thousands households. It is reported that the total electricity energy consumption of IDCs in the U.S. is about 75 billion kWh in 2010, which is about 2% of the total electricity consumption in the U.S. [2]. Consequently, the effective operation of IDCs in terms of energy efficiency attracts a significant attention to electric power utilities due to the potential impacts on power system security and adequacy.

In the meantime, IDCs operators could save a significant amount of money by making full utilization of time varying electricity prices and adjusting their energy consumption behaviors [3]. [4] proposed an optimal electric demand management strategy for large-scale IDCs distributed at different geographical locations and solved the energy efficiency problem via an MILP optimization model. [5] presented a bottom-up model to forecast future energy demand of data centers and showed the potential benefits by improving the energy efficiency. [6] designed a virtue machine-based strategy that could be applied to alleviate the hotspot in data centers and reduce the thermal energy consumption. [7] developed a thermal aware model to optimally schedule tasks and control temperatures of data centers for minimizing the total energy consumption. Impacts of various uncertainties on the IDC energy consumption have also been studied. [8] investigated the impacts of uncertain prices and workloads on IDC operation and employed the hedging scheme to handle risk issues. Since a higher temperature may increase the failure rate of IDC component clusters, including servers, hard drive disks (HDD), and routers, a data center was considered as a cyber-physical system [9]. It explored the energy consumption for both IT service and AC cooling operation, and evaluated the efficiency of the cyber-physical system control in data centers.

The cloud service is emerging across the world in recent years, which are composed of IDCs equipped with high-performance computing resources, high-capacity storage devices, and

high-speed networks that are shared among end users. Energy efficiency models for IDCs providing cloud services have been discussed in [10]. An electricity management framework considering shifting tasks among distributed IDCs in a cloud grid was addressed in [11].

Distributed IDCs, as large electricity consumers, are capable of adjusting their energy consumption behaviors by shifting tasks among IDCs. The economic benefit of the adjustment could be realized by participating into the DR program in the wholesale electricity markets [12]–[14]. The DR capacities would be used by Independent System Operators (ISOs) and/or Regional Transmission Owners (RTOs) to enhance operational security and economics of power systems. To this end, ISOs and RTOs are designing different DR programs based on their unique market rules [15]–[17].

While there is a growing concern on the electricity cost of cloud services and AC cooling operation, the environmental impacts of energy hungry IDCs have been less addressed [18]. While the commercial electricity rate in Texas is below the average electricity price in the U.S. [19], it has significantly higher carbon footprint as compared to the nation's average, which is measured by the normalized amount of carbon dioxide emissions per kilowatt hour electricity energy [20].

This paper proposes an electric DR management model for coordinating the operation of distributed large-scale IDCs, which optimizes the expected electricity payment minus the revenue from participating into the day-ahead DR program. If IDCs are operated individually without coordination, the electricity consumption requirement of dispersed IDCs may induce additional electricity transmission congestions and, in turn, degrade the performance and reliability of both IDCs and power systems. In addition, frequent ON/OFF switching would significantly increase a server's failure rate and deteriorate its performance. Major contributions of the paper include:

- 1) The DR capability of IDCs is rigorously formulated as the changes in the electricity consumption when shifting cloud service tasks among distributed IDCs in different time zones. The objective is to minimize the total electricity cost of IDCs, which includes costs of electricity for providing IT service, cooling, shifting cloud service tasks, the emission impacts of server cluster operations, and the revenue from participating in day-ahead DR programs. This study assumes that multiple servers in a cluster are configured identically and operated homogeneously in terms of the same frequency, while the required quality of service, reliability, and other specific constraints are satisfied.
- 2) The environmental impact of IDCs is taken into account by including the cost of carbon emission impacts into the IDC cost. Several states in different time zones with different carbon dioxide emission levels are selected to illustrate the effectiveness of proposed cloud service tasks shifting strategy in improving the environmental impacts.
- 3) The proposed model could help larger electricity customer like IDCs to determine their DR capabilities for enrolling into the day-ahead DR program and receiving revenues.

The rest of the paper is organized as follows. Section II describes the scheduling problem of distributed IDCs in smart grid environment. Section III proposes a scenario-based sto-

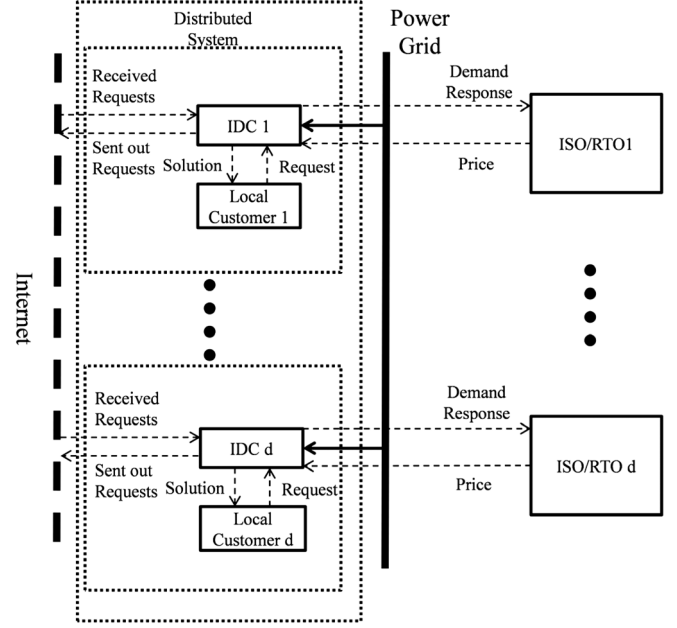


Fig. 1. Operation of distributed IDCs in smart grid environment.

chastic optimization model for the electric DR management of distributed large-scale IDCs. Illustrative case studies are presented in Section IV. The paper is concluded in Section V.

II. PROBLEM DESCRIPTION

Fig. 1 shows a graphical illustration of the proposed electric DR management strategy in the smart grid environment by coordinating distributed IDCs located in different time zones. Solid lines represent the power flow and dashed lines indicate the information/data flow. Smart meters will enable the bidirectional communication and control between the demand side and the electricity market in parallel with the power grid, which allow consumers to receive day-ahead electricity prices. IDCs are geographically distributed in different time zones, and they can work cooperatively with front-end routers and the wavelength division multiplexing (WDM) to support cloud services [10]. For instance, Google currently owns six data centers located in six states of U.S., including South Carolina, Iowa, Oregon, Georgia, North Carolina, and Oklahoma [21]. The six data centers are distributed in 3 time zones and different ISO control areas. In each electricity region, client service tasks from the local customer are firstly delivered to the local IDC. Then, it either stays locally or is forwarded to other IDCs through high-speed networks based on the task dispatch scheme. With the proposed model, IDC operators could make optimal decisions on dispatching cloud service tasks among distributed IDCs based on the day-head electricity price information and stochastic workloads arriving at IDCs, in order to minimize the expected electricity cost. The model can be used to determine the optimal DR capability of IDCs in each hour for participating into day-ahead DR programs while considering uncertain tasks of IDCs.

This paper targets a distributed IDC system in which individual IDCs may participate into different ISO markets as shown in Fig. 1. The case study in this paper investigates an IDC system including multiple IDCs located in TX, CA, and

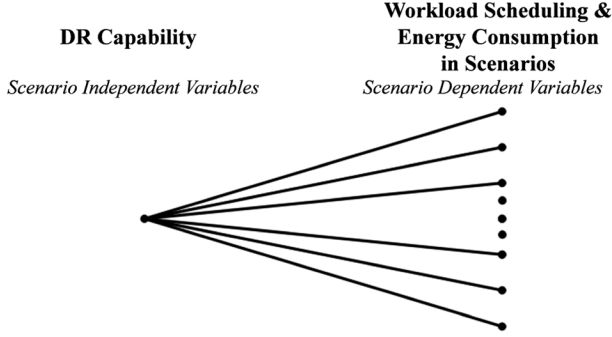


Fig. 2. Flow chart of the proposed stochastic optimization model for DR management.

NY. Thus, these IDCs participate into three electricity markets, including ERCOT, CAISO, and NYISO, which are facing with different emission levels and carbon emission tax rates. For instance, TX produces 0.62 tons/MWh CO₂ while NY produces 0.29 tons/MWh CO₂ in 2009 [20]. In addition, NY is about 2000 miles far from CA with a 3-hour time difference. Thus, different electricity prices caused by the east to west time zone effect provide distributed IDCs opportunities to shift tasks from the peak hour region to the off-peak hour region and save energy costs.

III. SCENARIO-BASED STOCHASTIC OPTIMIZATION MODEL

A. Objective of the Stochastic Optimization Model

The proposed model aims at assisting IDCs with decisions on how much capacity they can provide as DR in the day-ahead DR program, with the consideration of uncertain coming tasks. The proposed scenario-based stochastic optimization model will incorporate IDC workloads uncertainties and balance the tradeoff between the electricity bill payment and the DR revenue for distributed IDCs. The proposed model includes a base case for representing the forecasted workloads, and a set of scenarios generated via the MC simulation for representing workload uncertainties. It is assumed that the uncertainty of the number of cloud service tasks follows the normal distribution. Uncertain cloud service tasks for different IDCs may follow different distributions, which could be abstracted from history data and applied in the proposed model. As shown in Fig. 2, in the proposed scenario-based stochastic optimization model, the optimal DR schedule of each IDC DR_{dt} is scenario independent, which determines the total DR capability that an IDC can participate in the day-ahead DR programs. The optimal workload schedule of each IDC is scenario dependent for responding to workload uncertainties in each scenario. The optimized DR capability results DR_{dt} will determine DR capabilities of IDCs that they can participate in the day-ahead DR programs.

Computational requirements for solving scenario-based stochastic optimization models depend on the number of scenarios. Thus, an effective scenario reduction technique could be very essential. Scenario reduction can be performed by measuring certain probability metrics [22], [23], evaluating the impacts of individual scenarios on the objective function [24], [25], and matching specified statistical properties [26]. In this paper, scenario reduction is adapted to aggregate close scenarios and eliminate scenarios with very low probabilities, which would control

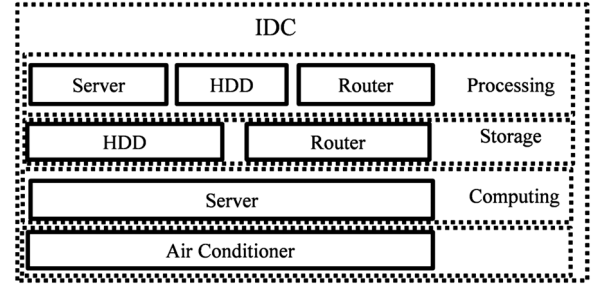


Fig. 3. Component clusters of an IDC.

the goodness-of-fit of approximation and improve the computational performance.

The proposed electric DR management problem for distributed IDCs is formulated as a mixed-integer linear programming (MILP) based stochastic optimization model. The objective (1) is to minimize the expected electricity bill for the base case and all scenarios throughout the entire scheduling horizon. It consists of expected electricity consumption costs for IT service, startup, cooling, and emission impacts of server cluster operations, as well as the transmitting task cost and the revenue from the DR provision. The sum of probabilities for the

base case and all scenarios is equal to one, i.e., $\beta^0 + \sum_{s=1}^{NS} \beta^s = 1$.

$$\min \left\{ \begin{aligned} &\beta^0 \cdot \sum_{t=1}^{NT} \sum_{d=1}^{ND} \left(C_{d,t}^{IT,0} + \sum_{i=1}^{NI_d} C_{i,d,t}^{SU,0} + C_{d,t}^{C,0} + C_{d,t}^{E,0} \right) \\ &+ \sum_{s=1}^{NS} \beta^s \cdot \sum_{t=1}^{NT} \sum_{d=1}^{ND} \left(C_{d,t}^{IT,s} + \sum_{i=1}^{NI_d} C_{i,d,t}^{SU,s} + C_{d,t}^{C,s} + C_{d,t}^{E,s} \right) \\ &+ \sum_{t=1}^{NT} \sum_{d=1}^{ND} C_{d,t}^T - RE \end{aligned} \right\} \quad (1)$$

Subject to various constraints as described below.

B. Constraints of IDCs

As shown in Fig. 3, an IDC is composed of a set of component clusters for providing IT service and the cooling operation. The former consists of server clusters, HDD clusters, and router clusters, and the latter contains AC clusters. Clusters in an IDC are divided into exclusive groups for handling different IT services, including processing (PR), storage (ST), and computing (CO) tasks. That is, each component cluster is specified to process a unique type of IT services. Similar to [4], it is assumed that all devices within a single cluster are operated homogeneously, i.e., at the same frequency and power level. The rest of the section is organized as follows. Subsection 1 describes the constraints for various cloud IT service tasks. Subsection 2 introduces the cooling control model for scheduling air conditioner clusters. Transmitting costs among distributed IDCs are presented in Subsection 3. The emission impact and the green IDC index PUE constraints are discussed in Subsection 4 to address the environment impacts. In the end, Subsection 5 formulates the revenue from participating in the day-ahead DR program.

1) Cost of IT Service:

Equation (2) describes the operation cost of an IDC for providing all IT services. The power consumption limits of an IDC

cluster include the maximum rated power and the minimum stand-by power levels (3).

$$C_{d,t}^{IT,s} = \sum_{i=1}^{NI_d} (c_{d,t} \cdot p_{i,d,t}^s) \quad (2)$$

$$P_{i,d}^{\min} I_{i,d,t}^s \leq p_{i,d,t}^s \leq P_{i,d}^{\max} I_{i,d,t}^s \quad (3)$$

In this study, three types of IT cloud services that are widely provided by IDCs across the world are considered, including processing, storage, and computing. These services are further classified into two categories based on their distinct characteristics: non-interruptible tasks and interruptible tasks. Non-interruptible means that a task cannot be stopped until it is finished. In this paper, processing and storage services are considered as non-interruptible tasks, and the computing services are considered as interruptible tasks. In addition, processing and storage services will be performed in the same hour when they are shifted among IDCs due to their quality of service (QoS) requirements, while the computing services could be deferred when they are shifted among IDCs, as long as they can be done within the pre-specified time requirements. In this paper, the size of a cloud task is measured in terms of the number of per unit tasks. The sizes of per unit processing and storage tasks are defined based on the information volume requirements of the tasks, i.e., 1.25 MB and 100 MB, respectively [4], [10]. The size of the per unit computing task is defined in terms of the CPU computing time requirement, i.e., 1 hour CPU time for per unit computing tasks. Thus, the size of a cloud task in terms of the number of per unit tasks can be calculated as the actual cloud task size over the per unit task size of the corresponding task type.

• Non-Interruptible Tasks

An IDC receives non-interruptible processing and storage tasks from local customers. The number of per unit non-interruptible tasks is limited by the capacity of a component cluster as shown in (4). To accommodate potential workload spikes, sufficient capacity margin in each hour is required for each IDC (5).

$$0 \leq \lambda_{i,d,t}^s \leq \lambda_{i,d}^{\max} \cdot I_{i,d,t}^s \quad (4)$$

$$\sum_{i=1}^{NI_d} (\lambda_{i,d}^{\max} - \lambda_{i,d,t}^s) \geq CM_d \quad (5)$$

Two IDC component clusters, HDD clusters and router clusters, are called for providing storage services, such as upload or download actions. In this paper, power consumptions of HDD clusters and router clusters are formulated as linear functions of the task rate (6), which has been widely used in literature for representing the relationship between the power consumption and the utilization of component clusters [5]. Other non-linear relationships described in [9] and [28] could also be used, which can be piecewise linearized and incorporated into the proposed MILP model.

$$p_{i,d,t}^s = \frac{\lambda_{i,d,t}^s}{\lambda_{i,d}^{\max}} \cdot P_{i,d}^{\max}, \quad i \in \mathbf{T}_{\text{NIT}} \quad (6)$$

A processing service will involve server, HDD, and router component clusters. The power consumption of the HDD and

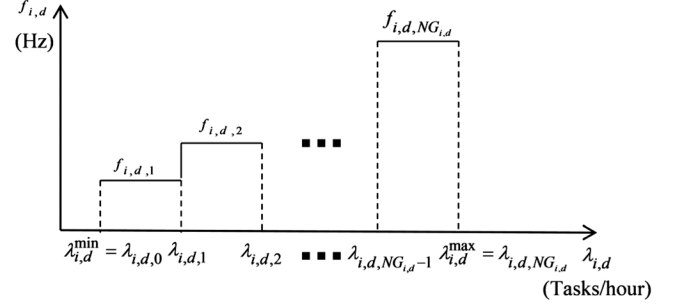


Fig. 4. Discrete frequency levels for a server cluster handling processing tasks.

router clusters is the same as (6). Server clusters could be operated under a set of discrete CPU frequency levels as shown in Fig. 4, and a certain level will be selected depending on the upcoming stochastic workloads and the quality of service. Equation (7) calculates the power consumption of a server cluster, which includes the power consumption related to the hardware and the server utilization. Equations (8)–(10) represent that a server cluster will adopt a frequency from a set of predefined discrete levels based on the number of tasks to be processed. The upper limit of the CPU utilization (11) prevents the CPU from saturation. Equation (12) uses a M/M/1 queue model to represent the QoS requirement on each server [4].

$$p_{i,d,t}^s = B_{i,d} \cdot I_{i,d,t}^s + A_{i,d} \cdot (f_{i,d,t}^s)^3 \cdot \frac{\lambda_{i,d,t}^s}{\lambda_{i,d}^{\max}}, \quad i \in \mathbf{T}_{\text{NIT}} \quad (7)$$

$$\sum_{g=1}^{NG_{i,d}} \lambda_{i,d,g-1} \cdot \delta_{i,d,t,g}^s \leq \lambda_{i,d,t}^s \leq \sum_{g=1}^{NG_{i,d}} \lambda_{i,d,g} \cdot \delta_{i,d,t,g}^s \quad (8)$$

$$\sum_{g=1}^{NG_{i,d}} \delta_{i,d,t,g}^s = I_{i,d,t}^s \quad (9)$$

$$f_{i,d,t}^s = \sum_{g=1}^{NG_{i,d}} f_{i,d,g} \cdot \delta_{i,d,t,g}^s \quad (10)$$

$$U_{i,d,t}^s = \frac{\lambda_{i,d,t}^s}{\lambda_{i,d}^{\max}}, \quad 0 \leq U_{i,d,t}^s \leq U_{i,d}^{\max} \quad (11)$$

$$I_{i,d,t}^s \leq D_{i,d} \cdot (\lambda_{i,d}^{\max} - \lambda_{i,d,t}^s) \quad (12)$$

• Interruptible Tasks

A computing service is simulated as an interruptible task in this paper. This task only employs server clusters and the power consumption is formulated in (13). Some computing tasks have pre-specified total execution time measured in terms of number of computing hours on a single cluster capability. That is, tasks assigned to an IDC in scenario s would use H_d^s hours to be completed by single server cluster and the workload balance with several clusters is shown in (14). Equation (15) enforces all computing tasks will be completed.

$$p_{i,d,t}^s = P_{i,d}^{\max} \cdot I_{i,d,t}^s, \quad i \in \mathbf{T}_{\text{IT}} \quad (13)$$

$$\sum_{t=1}^{NT} \sum_{i=1}^{NI_d} I_{i,d,t}^s = H_d^s - R_d^S + R_d^R \quad (14)$$

$$\sum_{t=1}^{NT} \sum_{d=1}^{ND} \sum_{i=1}^{NI_d} I_{i,d,t}^s = \sum_{d=1}^{ND} H_d^s \quad (15)$$

The startup cost is calculated as (16), while (17) describes the relationship among startup/shutdown indicator and ON/OFF status of a server cluster [4]. The startup indicator is equal to 1 if a server cluster is invoked, otherwise it remains 0. Similarly, the shutdown indicator is equal to 1 if a server cluster is shut down from running status and 0 otherwise. Both startup and shutdown indicators are binary variables. In order to reduce the number of binary variables and improve the model efficiency, startup and shutdown indicators could be equivalently modeled as continuous variables with additional constraints (18), (19).

$$C_{i,d,t}^{SU,s} = SU_{i,d} y_{i,d,t}^s \quad (16)$$

$$y_{i,d,t}^s - z_{i,d,t}^s = I_{i,d,t}^s - I_{i,d,t-1}^s \quad (17)$$

$$y_{i,d,t}^s \leq 1 - I_{i,d,t-1}^s, \quad z_{i,d,t}^s \leq I_{i,d,t-1}^s \quad (18)$$

$$0 \leq y_{i,d,t}^s \leq 1, \quad 0 \leq z_{i,d,t}^s \leq 1 \quad (19)$$

2) Cost of Cooling:

ACs are used to maintain the indoor temperature where component clusters are located at, since IDC component clusters are recommended to be operated in a predefined temperature region $[T^{\min}, T^{\max}]$ for reducing the failure rates. The cost associated with AC operation is described in (20). The lower and upper capacity limitations of each AC cluster is stated in (21). Equation (22) describes the relationship between the AC supplied temperature and the indoor temperature, which depends on the energy consumption of all IDC clusters. According to the law of energy conservation and the fact that almost all power drawn by an IDC component cluster is dissipated as heat [7], the relationship among the power consumption of all component clusters, the AC supplied temperature, and the indoor temperature can be represented as (23). That is, the power consumption of cluster i will raise the air temperature from AC supplied temperature $T_{d,t}^{\text{sup},s}$ to indoor temperature $T_{d,t}^{\text{out},s}$. Equation (24) describes that an AC cluster will be switched off when the AC supplied temperature is cooled down to the level of T^{\min} . Equations (25), (26) depict that if the AC supplied temperature is higher than T^{\max} , an AC cluster will run at its maximum power in order to approach the required temperature level as quick as possible [29].

$$C_{d,t}^{C,s} = c_{d,t} \cdot p_{d,t}^{C,s} \quad (20)$$

$$PC^{\min} I_{d,t}^{C,s} \leq p_{d,t}^{C,s} \leq PC^{\max} I_{d,t}^{C,s} \quad (21)$$

$$T_{d,t}^{\text{sup},s} = \varepsilon \cdot T_{d,t-1}^{\text{sup},s} + (1 - \varepsilon) \cdot (T_{n,d,t}^{\text{out},s} - \eta_{ac} \cdot p_{d,t}^{C,s}) \quad (22)$$

$$C_p \cdot f_d \cdot \rho \cdot T_{d,t}^{\text{out},s} = C_p \cdot f_d \cdot \rho \cdot T_{d,t}^{\text{sup},s} + \sum_{i=1}^{N I_d} p_{i,d,t}^s \quad (23)$$

$$T_{d,t}^{\text{sup},s} - T^{\min} \geq M \cdot (I_{d,t}^s - 1) \quad (24)$$

$$M \cdot (u_{d,t}^s - 1) \leq T^{\max} - T_{d,t}^{\text{sup},s} \leq M \cdot u_{d,t}^s \quad (25)$$

$$PC^{\max} \cdot (1 - u_{d,t}^s) \leq p_{d,t}^{C,s} \quad (26)$$

3) Cost of Shifting Service Among Distributed IDCs:

IT tasks can be shifted among distributed IDCs via the advanced optical fiber technology within a very short period of time. Thus, it is reasonable to assume that tasks shifted out from one IDC can arrive at other IDCs in the same hour. As the energy consumption through optical transceivers and optical cables is relatively small as compared to that of the IDC operation [30],

the transmitting cost will include only the optical fiber interface energy consumption. The electrical energy for shifting IT cloud tasks is mainly consumed by the WDM operation, which is represented in (27), (28). Equation (29) describes the sending status of an IDC with its optimal cable connections. The transmitting capacity of each optical cable is shown in (30), (31). (32) shows that an IDC cannot receive and send tasks through the same transmission line at the same time. However, an IDC can shift out and receive tasks simultaneously via different cables.

$$C_{d,t}^T = c_{d,t} \cdot p_{d,t}^W \quad (27)$$

$$p_{d,t}^W = \sum_{l=1}^{N L_d} P T_l \cdot \mu_{l,d,t} \quad (28)$$

$$0 \leq \sum_{l=1}^{N L_d} \mu_{l,d,t} \leq N L_d \cdot I_{d,t}^W \quad (29)$$

$$R_{l,d,t}^S \leq F L_l^{\max} \mu_{l,d,t} \quad (30)$$

$$R_{l,d,t}^R \leq F L_l^{\max} \gamma_{l,d,t} \quad (31)$$

$$\gamma_{l,d,t} + \mu_{l,d,t} \leq 1 \quad (32)$$

4) Cost of Carbon Emission Impacts:

The electricity for the operation of distributed IDCs may come from various energy sources under distinct carbon emission levels. Traditional thermal generating units have higher carbon footprint and lower prices than renewable wind and hydro units [19]. Thus, it may bring potential adverse consequences to environment if a heavily loaded IDC is located in an area where electricity is mostly generated by thermal units. If the carbon tax has been implemented on generating units, the cost of CO₂ will be passed on to electricity consumers through locational marginal prices (LMPs). However, currently in U.S., only Boulder in Colorado and Bay Area Air Quality Management District in California have implemented carbon tax [32], and the general policy is still in developing but not popular yet. Considering that the carbon tax is not fully implemented, LMPs in current ISOs do not fully reflect the emission cost. For instance, for the three ISOs (i.e., NYISO, ERCOT, and CAISO) considered in numerical case studies, ERCOT has the highest emission level and the lowest LMPs. Thus, similar as the idea implemented in [18], carbon rates are employed in this paper to evaluate the environmental impacts of large-scale distributed IDCs at different ISOs. Since carbon emission rates are not popular in most states in US, the current carbon emission rate in Colorado is employed as the baseline for calculating equivalent carbon emission rates in all other states. For instance, as reported in [18], the emission level in Wyoming is 1.8 times as high as that in Colorado, and Colorado currently sets its carbon emission rate as 0.49 cents/kWh for residential customers. Thus, the equivalent carbon emission rate for residential customers in Wyoming can be calculated as $0.49 \times 1.8 = 0.88$ cents/kWh. Following the similar idea in [18], given that the average carbon dioxide emission level of the power sector in Texas is about 0.62 tons/MWh and is about 0.87 tons/MWh in Colorado [20], and the carbon tax for industry customer is 0.03 cents/kWh in Colorado Boulder [36], the equivalent carbon emission rate in Texas used in this paper is calculated as $0.03 \times 0.62/0.87 = 0.02$ cents/kWh. Since IDCs under study are distributed in different states, the

inclusion of carbon rates could better emphasize the different composition of energy for electricity generation in different ISOs with variant emission levels, and evaluate the environmental impacts of distributed IDCs.

In this paper, the cost of carbon emission impacts is included in the objective function (1) for considering the environmental impact of IDCs. For each IDC, the emission impact is calculated using the average carbon dioxide emission rate ϕ_d as shown in (33), which represents the average cost of carbon emission impacts of 1 kWh electric energy in a region. In addition, the power utilization effectiveness (PUE) [31] is adopted to measure the energy utilization efficiency of data centers, which is calculated as the ratio of the total energy consumed by data centers over the energy used for IT services in (34).

$$C_{d,t}^{E,s} = \phi_d \cdot \left(\sum_{i=1}^{NI_d} p_{i,d,t}^s + p_{d,t}^{C,s} + p_{d,t}^W \right) \quad (33)$$

$$\frac{\sum_{i=1}^{NI_d} p_{i,d,t}^s + p_{d,t}^{C,s} + p_{d,t}^W}{\sum_{i=1}^{NI_d} p_{i,d,t}^s + p_{d,t}^W} \leq PUE_d \quad (34)$$

5) Revenue From the Demand Response Program:

In this paper, the DR provision means that if an IDC can shift tasks out without compromising the quality of service of IDCs or violating transmitting capacities, the reduced energy consumption at certain IDC locations could be considered as the DR provision capacity that can bid into day-ahead DR programs and obtain revenues. The DR provision capabilities of IDCs can be used to provide ancillary reserves, as an Up ramp resource [35]. It is a key issue to measure the amount of DR that could be offered by IDCs, which is the energy consumption reduction deviated from the baseline level of IDCs. However, unlike the metered energy consumption, the DR capability is difficult to quantify as there is no way to directly measure the baseline power consumption and the energy consumption reduction [34]. [10] discussed that the power consumption requirement of a cloud task is relative to the task size, e.g., in the unit of MBs for storage tasks. In this paper, the size of cloud tasks is measured in terms of the number of per unit tasks, and the average per unit task energy consumptions are obtained from [4] and [10] as 1.8×10^{-6} kWh, 2.1×10^{-4} kWh, and 106.5 kWh for processing, storage, and computing tasks, respectively. Since each component cluster is specified to process a unique type of IT service tasks as shown in Fig. 3, the energy consumption of a per-unit cloud service task on cluster i E_i is the average per unit task energy consumption value of the corresponding type of tasks. Thus, the DR capability of an IDC can be quantified as the potential energy saving by shifting could service tasks, which is calculated by E_i and the number of cloud service tasks shifted as shown in (37). If the IDC is sending out net tasks, the associated DR capability would be the corresponding energy saving; otherwise, it is zero. The relationship is expressed in (35), (36) for non-interruptible tasks and (37), (38) for interruptible tasks. As non-interruptible tasks need to be finished within the same hour, DR provided by shifting non-interruptible tasks is time varying based on the number of tasks shifted in each hour. On the other hand, interruptible tasks could be delayed and we focus on the number of shifted computing tasks for

the entire day. In this paper, workload uncertain is simulated via scenarios with different numbers of interruptible tasks received by individual IDCs, and the total number of tasks received by all IDCs is constant. Equation (39) enforces that all non-interruptible tasks will be completed in the same hour by IDCs. Non-interruptible task balance for each cluster is stated in (40) considering the out-shifting and in-shifting tasks. On the other hand, interruptible computing tasks take a very small proportion and their formation content is quite low [4]. Thus, their transmitting cost for this type of task is reasonable neglected. DR capacity limit is given in (41). The revenue of IDCs from the DR provision is calculated in (42), in which the DR reward rates are modified from [34] to measure the revenue for DR participants.

$$0 \leq DR_{d,t} - \sum_i E_i \cdot \left(\sum_{l=1}^{NL_d} R_{l,i,d,t}^S - \sum_{l=1}^{NL_d} R_{l,i,d,t}^R \right) \leq M \cdot (1 - I_{d,t}^{DR}) \quad (35)$$

$$0 \leq DR_{d,t} \leq M \cdot I_{d,t}^{DR} \quad (36)$$

$$0 \leq DR_d - \sum_i E_i \cdot (R_{i,d}^S - R_{i,d}^R) \leq M \cdot (1 - I_d^{DR}) \quad (37)$$

$$0 \leq DR_d \leq M \cdot I_d^{DR} \quad (38)$$

$$\sum_{d=1}^{ND} \lambda_{d,t}^s = \sum_i R_{i,t} \quad (39)$$

$$\lambda_{d,t}^s = \sum_i \left(R_{i,d,t}^s - \sum_{l=1}^{NL_d} R_{l,i,d,t}^S + \sum_{l=1}^{NL_d} R_{l,i,d,t}^R \right) \quad (40)$$

$$DR_{d,t} \leq q_d^{\max} I_{d,t}^W \quad (41)$$

$$RE = \sum_{d=1}^{ND} b_d \left(\sum_{t=1}^{NT} DR_{d,t} + DR_d \right) \quad (42)$$

IV. CASE STUDIES

In this section, three types of typical component clusters, including server, HDD, and router, are utilized to study the proposed optimal electric DR management for distributed IDCs via the proposed stochastic optimization approach. The component cluster data are presented in Tables I, II, which are modified based on the information from [4], [7], [10], and [29]. The last row in Table I indicates the number of components for each cluster type. For instance, each HDD cluster that is assigned to perform PR tasks is composed of 20 HDDs, and each HDD cluster that is assigned to perform ST tasks is composed of 500 HDDs. Each computing server cluster is composed of 3000 computing servers. Threshold value for the processing server is set as 2.7 million per unit tasks per hour as shown in Table II. The boundary of the AC supplied temperature is $[68^\circ\text{F}, 72^\circ\text{F}]$, and the initial AC supplied temperature and the indoor temperature are both 70°F . The information for processing tasks and storage tasks are modified from [4], [10]. The sizes of per unit processing and storage tasks are defined based on the information volume requirements of the tasks, i.e., 1.25 MB and 100 MB, respectively [4], [10]. The size of the per unit computing task is defined in terms of the required CPU computing time, i.e., 1 hour CPU time for per unit computing tasks. The average per unit task energy consumptions of the three types of tasks

TABLE I
DAILY OPERATION DATA FOR IDC COMPONENTS

Terms	HDD (PR/ST)	Router (PR/ST)	AC	WDM
Maximum Power Level (kW)	4.9	10.9	3.5	0.136 /500mile
Capacity	604.8Tb	640Gb/s	N/A	40Gb/s
Number of components for each cluster type	20/500	10/20	1420	4

TABLE II
SERVER CONFIGURATION FOR PROCESSING TASK

	Frequency (GHz)	Capacity (Million/h)	Number per cluster	B (kW)	A (kW/GHz ³)	Utilization limit
High	3.4	3.06	5500	68	4.485	0.9
Low	3.0	2.7	5500	53	3.206	0.9

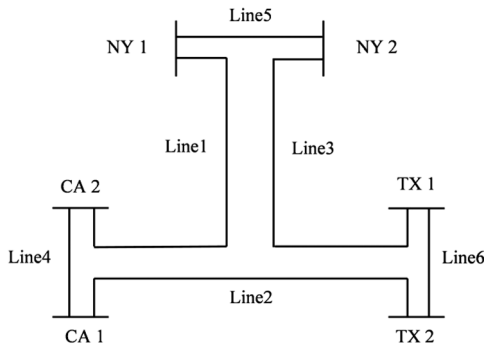


Fig. 5. Distributed large scale IDCs.

TABLE III
DATA FOR THE SIX DISTRIBUTED IDCs

Terms	CA 1	CA 2	NY 1	NY 2	TX 1	TX 2
Processing Capacity (Million/h)	442	442	221	664	221	221
Storage Capacity (Million/h)	54.1	54.1	27	81.2	27	27
PUE	1.3	1.5	1.2	1.4	1.8	2
Number of HDD Clusters (PR/ST)	2/2	2/2	1/1	3/3	1/1	1/1
Number of Router Clusters (PR/ST)	2/2	2/2	1/1	3/3	1/1	1/1
Number of Server Clusters (PR/CO)	2/2	2/2	1/1	3/3	1/1	1/1

are 1.8×10^{-6} kWh, 2.1×10^{-4} kWh, and 106.5 kWh [4], [10], which are used to calculate the electric DR capability of IDCs. The required execution time of computing tasks is much higher than those of processing and storage tasks. Thus, the energy consumption of a per unit computing task is several orders of magnitude higher than the others. Day-ahead LMPs of three certain locations in CAISO [15], NYISO [16], and ERCOT [17] on September 1st, 2012 are used in numerical case studies to show the impact of dispersed IDCs at different time zones. All case studies are conducted utilizing CPLEX 12.4.0 on an Intel Core-i7 3.4-GHz personal computer.

The six IDCs are distributed in three states as depicted in Fig. 5. IDCs are connected via 6 optical cables, which allow them to shift tasks to other IDCs quickly for the purpose of cloud computing. Table III provides information for individual IDCs,

including capacities for handling PR and ST IT service tasks, the PUE upper limit for measuring the power utilization efficiency in terms of the energy consumption ratio between IT service and AC operation, and numbers of different types of clusters. For instance, the second column shows that CA1 can handle 442 million per unit processing tasks and 54.1 million per unit storage tasks in any single hour, and the total energy consumption over the IT service energy consumption needs to be no larger than 1.3. In addition, CA1 has 2 HDD clusters, 2 router clusters, and 2 server clusters to handle processing tasks, 2 HDD clusters and 2 router clusters to deal with storage tasks, and 2 server clusters for computation tasks. The DR upper limits are all set as 15 MW while capacity margins are all set as 10000 per unit tasks/hour for processing and storage tasks. The distances of CA-NY, CA-TX, and NY-TX are set as 3000 miles, 2000 miles, and 1500 miles, respectively, and distances of IDCs within the state are all set as 500 miles. The initial durations for computing tasks are sampled from the set {15, 10, 9, 24, 5, 7} (hours). The carbon rates for industry customers in CA, NY, and TX are set as 0.0096 cents/kWh, 0.01 cents/kWh, and 0.0215 cents/kWh, respectively [20], [36].

In this study, the probability of base case is calculated based on the 95% confidence level of the normal distribution, i.e., 0.68. In addition, 1000 scenarios are generated by the Monte Carlo method, which follow the same normal distribution and cover the remaining 5% confidence interval for simulating hourly workload uncertainties. They are further reduced to 50 scenarios via the scenario reduction technique and the corresponding probabilities are derived by the scenario reduction [27] using GAMS [37].

The following three cases are studied:

Case 0: Six large-scale distributed IDCs are operated uncoordinatedly and do not provide DR.

Case 1: Six large-scale distributed IDCs are operated coordinately and provide DR.

Case 2: The sensitivity analysis with respect to different DR capability upper limits of IDCs.

Case 0: Six large-scale distributed IDCs are operated uncoordinatedly and do not provide DR

Case 0 is used as the base case for comparing with Case 1 and showing the benefits of the proposed model. In this case, six large-scale distributed IDCs are operated uncoordinatedly for optimizing their individual costs. That is, in Case 0, IDCs do not provide DR and only constraints (2)–(26) and (33), (34) are considered. The total cost is $\$1.21 \times 10^4$ and the energy consumption is 379.12 MWh.

Case 1: Six large-scale distributed IDCs are operated coordinately and provide DR

This case evaluates the proposed optimal DR management of IDCs. The input data are the same as Case 0. The result indicates the total cost is reduced to $\$1.06 \times 10^4$, as compared to $\$1.21 \times 10^4$ in Case 0. Thus, a save of 12.4% can be obtained by providing DR.

The final optimal non-interruptible task shifting results among all six IDCs are listed in Table IV. In each cell, the first value indicates the number of tasks an IDC shifts out while the second value represents the number of tasks received. It is observed that IDCs with larger cluster capacities, such as CA1 and NY2, receives significant amounts of tasks from

TABLE IV
NON-INTERRUPTIBLE TASKS SHIFTING AMONG SIX IDCs

Tasks (10^7)	CA 1	CA 2	NY 1	NY 2	TX 1	TX 2
Line1	0/0	7.96/0	0/7.96	0/0	0/0	0/0
Line2	5.88/8.23	0/0	0/0	0/0	0/0	8.23/5.88
Line3	0/0	0/0	0/0	1.15/14.3	14.3/1.15	0/0
Line4	3.75/15.0	15.0/3.75	0/0	0/0	0/0	0/0
Line5	0/0	0/0	9.8/6.94	6.94/9.8	0/0	0/0
Line6	0/0	0/0	0/0	0/0	12.9/0	0/12.9
Balance	0/13.6	19.21/0	0/5.1	0/16.01	26.05/0	0/10.55

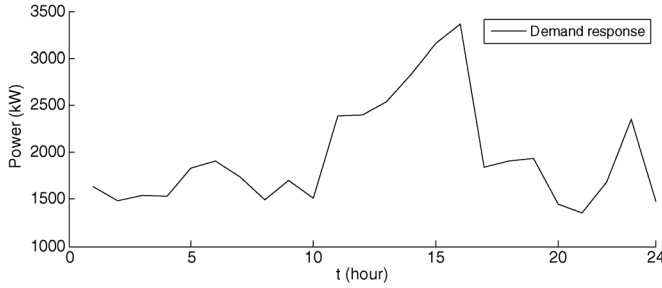


Fig. 6. DR (energy reduction) for IDCs by shifting non-interruptible tasks.

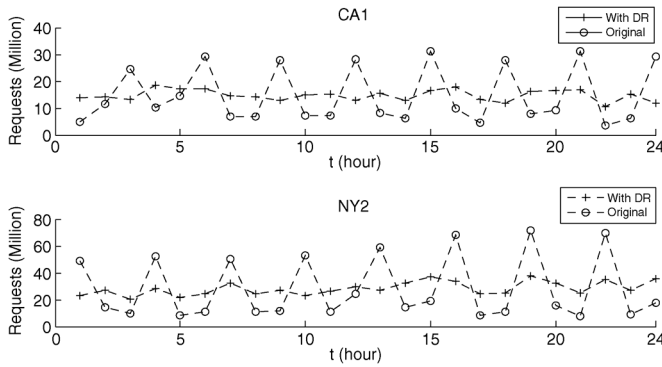


Fig. 7. Non-interruptible task shifts of CA1 and NY2.

other IDCs, while the most significant shifting is from smaller scaled IDCs, such as CA2 and TX1. Furthermore, differences on electricity price, tasks workloads, and DR reward rates will also contribute to shifting decisions. The number of tasks processed in high electricity price region TX1 is reduced while the CA1 with low electricity prices processes more tasks.

The DR provided by the six IDCs is depicted in Fig. 6. It is shown that non-interruptible tasks are mostly shifted out from peak workload hours (in hours 11–17) by providing DR. Fig. 7 illustrates the final non-interruptible task shifting decisions of CA1 and NY2 for providing DR across 24 hours. It shows the original and the final workloads before and after the implementation of DR. It indicates that the task shifts will fill up the valley of workload capacity profile by receiving tasks and shed workload peaks by sending out tasks, which demonstrates that DR will shave the peak tasks workloads and further flatten the corresponding load profile.

Fig. 8 illustrates the final interruptible task shifting decision for providing DR and the revenues for the entire day. It shows that tasks will be shifted out of CA2, NY2, and TX1 to CA1. The main reason is that CA has lower DR reward rates and lower electricity prices, which would drive the tasks to be shifted from other

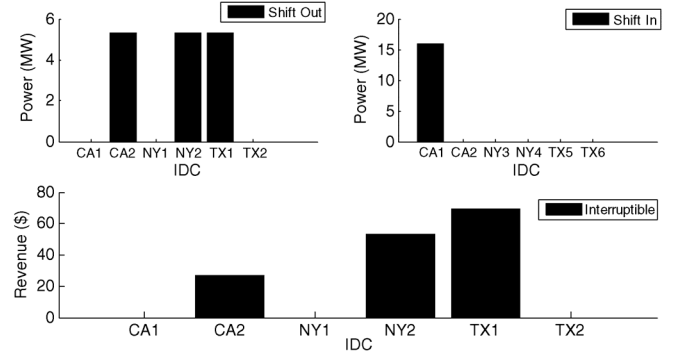


Fig. 8. Revenue and shifted interruptible tasks with DR among IDCs.

TABLE V
ENERGY CONSUMPTION OF THE SIX IDCs IN CASES 0 AND 1

Energy (MWh)	CA 1	CA 2	NY 1	NY 2	TX 1	TX 2
Case0 Operation	61.61	60.72	59.09	74.04	54.91	68.75
Case1 Operation	89.7	43.41	61.98	75.51	33.85	74.67
Transmitting	0.09	0.18	0.02	0.04	0.14	0.08

IDCs. The system topology constraints, such as information flow congestions, differences of workloads and electricity prices in different time zones, and the optical fiber network topology also contribute to the shifting of tasks. Fig. 8 also shows the revenue of IDCs by participating into the day-ahead DR program. Although CA2 shifts out an identical number of tasks, its revenue rewarded is smaller as compared to other IDCs. NY1 and TX2 have zero DR revenues as they do not shift out or receive interruptible tasks. CA1 obtains zero revenue as it only receives interruptible tasks but does not shift any out.

The savings in CO₂ is achieved by shifting cloud IT service tasks among distributed IDCs and, in turn, the electricity consumption from areas with high carbon footprint to those with low emission levels. The carbon rate is considered as part of the operation cost of IDCs, which is minimized in the objective function for emphasizing the impacts on environment and achieving carbon reduction [33]. The savings in CO₂ can be quantified by Independent System Operators of power systems with emission curves of generating units. This paper stands from IDCs' viewpoint, which usually do not have information on emission curves of generating units for calculating the amount of CO₂ incurred at the generation side. Alternatively, the savings in CO₂ are approximated by the average carbon dioxide emissions rate of each ISO. The average carbon dioxide emission rates of California, New York, and Texas are 0.28 tons per MWh, 0.291 tons per MWh, and 0.62 tons per MWh, respectively [20]. Thus, the carbon emission rate of Texas is set higher than those of NY and CA, in order to take the benefit of lower emission levels in NY and CA and alleviate the environmental impacts in US. That is, in terms of the overall nationwide situation, coordinated operation among distributed IDCs would be more environmentally friendly than IDCs operating independently on their own. Furthermore, a higher carbon emission rate would encourage more investments in renewable and environmentally friendly energy such as wind generation, which could consequently reduce the per MWh emission level of the power section and, in turn, the carbon emission rates step by step. Table V shows the energy consumption of the six IDCs

in Case 0 and Case 1. Thus, the savings of 4.92 tons CO₂ is achieved in Case 1 as compared to Case 0.

The energy consumption may increase due to the coordinated scheduling among distributed large-scale IDCs. The reason is that the objective is to minimize the total operation cost of distributed IDCs by shifting cloud IT service tasks among distributed IDCs, but not the total energy consumption. Thus, the energy consumption may increase due to the coordination among distributed large-scale IDCs. Case 1 reduces the total cost by $\$0.15 \times 10^4$ (i.e. $1.21 \times 10^4 - 1.06 \times 10^4$), but increases the total energy consumption by 0.54 MWh (i.e. $379.67 - 379.13$) as compared to Case 0, as shown in Table V. It is noted that in Case 1, the energy consumption for transmitting tasks is 0.55 MWh and is 379.12 MWh for the other operation needs of IDCs, which is smaller than 379.13 MWh in Case 0. Thus, the increase of the total energy consumption in Case 1 is mainly caused by the additional energy request from shifting tasks, and the energy consumption for the operation needs of IDCs is reduced as compared to Case 0.

Case 2: The sensitivity analysis with respect to different DR capability upper limits of IDCs

This case performs a sensitivity analysis for exploring the impacts of different pre-specified DR upper limits on the DR management efficiency of IDCs. In this paper, efficiency of task shifting described the efficiency of an IDC for shifting its task out to other IDCs while not violating operation constraints, which would reflect the demand response management efficiency of an IDC. A normalization index IE (43) calculates the ratio of shifted tasks over the total number of arrived tasks, which would describe the efficiency of tasks shifting among distributed IDCs. Thus, a bigger IE value indicates that a higher percentage of original tasks can be shifted, and thus a higher efficiency of tasks shifting. Usually a bigger IDC would have a larger capability of tasks shifting in terms of the actual number of shifted tasks, but not necessarily a higher efficiency. For instance, given that 1000 tasks originally arrive at a larger IDC1 and 100 tasks arrive at a smaller IDC2. 200 tasks from IDC1 and 40 tasks from IDC2 are shifted out to other IDCs. Thus, IE for IDC1 is 0.2 (i.e., $200/1000$) and is 0.4 (i.e., $40/100$) for IDC2, which indicates that although IDC1 shifts out 160 more tasks than IDC2, IDC2 has a higher efficiency of task shifting in terms of the ratio of shifted tasks over the total number of arrived tasks. In case 2, the total number of tasks arrived at the IDC system is fixed. The Sensitivity of the DR upper limit on IE , the total electricity cost, and the DR revenue are explored, which indicates that a higher IE value would correspond to a higher DR management efficiency of IDCs.

$$IE = \frac{\sum_i \sum_{d=1}^{ND} \left(\sum_{t=1}^{NT} R_{i,d,t}^R + R_{i,d}^R \right)}{\sum_i \left(\sum_{t=1}^{NT} R_{i,t} + \sum_{d=1}^{ND} R_{i,d} \right)} \quad (43)$$

The impacts of DR upper limits on scheduled daily costs, shifting efficiency, DR provision, and DR revenues are depicted in Fig. 9. It is shown that the scheduling efficiency, DR quantity, and DR revenues are all increased while the scheduled daily cost drops, till the DR upper limit is increased up to a certain level. The reason is that an IDC may obtain more revenue if

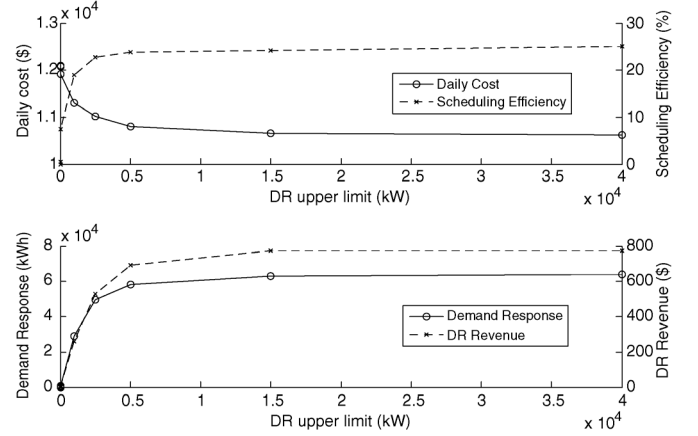


Fig. 9. Sensitivity analysis.

they are allowed to shift more tasks while not compromising the quality of service of IDCs or violating transmitting capacities. It is reasonable that larger scale IDCs would have more flexibility when processing the same number of tasks while satisfying all physical limits. In addition, when the DR upper limit reaches 15 MW, all values remain unchanged. This indicates that the DR capability of an IDC is restricted by the physical operation limits of its clusters. Thus, an IDC may not want to invest in excessive transmitting equipment, e.g., WDM, to support DR that may not realize beyond the upper limit, which is around 15 MW in this case.

V. CONCLUSIONS

With the growth of distributed large-scale IDCs throughout the world, the impacts of IDCs on power systems become more apparent. Economic incentives and compensations for peak load reduction provide motivations for large electricity consumers like IDCs to optimally manage their electricity consumption. Due to the potential economic benefits that the day-ahead electricity price and the DR program could bring to the demand side, IDCs could optimally adjust the energy consumption by shifting IT cloud service tasks among distributed IDCs and providing DR capabilities.

This paper proposed a stochastic optimization model to evaluate the electric DR management for distributed large-scale IDCs, which determines the optimal hourly DR capability of IDCs while considering uncertain cloud service tasks to IDCs and operation constraints of individual component server clusters. The thermal management formulation is included to calculate the energy consumption of cooling system, and the environmental factors are also included via the consideration of carbon emission rates. Numerical case studies and sensitivity analysis show that the proposed management scheme could reduce the electricity costs by coordinating the operation of distributed IDCs and providing DR. Furthermore, the DR capabilities of IDCs are restricted by their physical operation limits. The proposed DR management solution could help IDC operators make optimal decisions to participate in the day-ahead DR program by effectively shifting their electricity demand among distributed IDCs while achieving cost savings. The model will assist IDCs for determining their DR capabilities when participating in day-ahead DR programs, but not

intending to derive fast response real-time schedules. As shown in numerical case studies, the average computational time of about 30 minutes is sufficient for performing the proposed day-ahead schedule. Efficient decomposition techniques, such as Benders decomposition [38], may be used for solving the proposed large-scale NP-hard MILP problem, which will be explored to study fast response real-time schedule of IDCs in the future work.

REFERENCES

- [1] EPA, Report to Congress on Server and Data Center Energy Efficiency Aug. 2007.
- [2] J. G. Koomey, *Growth in Data Center Electricity Use 2005 to 2010*. Oakland, CA: Analytics Press, Aug. 2010.
- [3] A. Qureshi, R. Weber, H. Balakrishnan, J. Guttag, and B. Maggs, "Cutting the electric bill for Internet-scale system," in *ACM SIGCOMM Conference*, Aug. 2009.
- [4] J. Li, Z. Li, K. Ren, and X. Liu, "Towards optimal electric demand management for Internet data centers," *IEEE Trans. Smart Grid*, vol. 3, no. 1, pp. 183–192, Mar. 2012.
- [5] E. R. Masanet, R. E. Brown, A. Shehabi, J. G. Koomey, and B. Nordman, "Estimating the energy use and efficiency potential of U.S. data centers," *Proceedings of the IEEE*, vol. 99, no. 8, pp. 1440–1453, Aug. 2011.
- [6] I. Rodero, H. Viswanathan, E. K. Lee, M. Gamell, D. Pompili, and M. Parashar, "Energy-efficient thermal-aware autonomic management of virtualized HPC cloud infrastructure," *Journal of Grid Computing*, vol. 10, no. 3, pp. 447–473, Sep. 2012.
- [7] Q. Tang, S. K. S. Gupta, D. Stanzione, and P. Cayton, "Thermal-aware task scheduling to minimize energy usage of blade server based datacenters," in *Proc. Second IEEE Int'l Symp. Dependable, Autonomic and Secure Computing (DASC '06)*, Oct. 2006.
- [8] L. Rao, X. Liu, L. Xie, and P. Zhang, "Hedging against uncertainty: A tale of Internet data center operations under smart grid environment," *IEEE Trans. Smart Grid*, vol. 2, no. 3, pp. 555–563, Sept. 2011.
- [9] L. Parolini, B. Sinopoli, B. H. Krogh, and Z. Wang, "A cyber-physical systems approach to data center modeling and control for energy efficiency," *Proc. of the IEEE*, vol. 100, no. 1, pp. 254–268, Jan. 2012.
- [10] J. Baliga, R. W. A. Ayre, K. Hinton, and R. S. Tucker, "Green cloud computing: Balancing energy in processing, storage, and transport," *Proceedings of the IEEE*, vol. 99, no. 1, pp. 149–167, Jan. 2011.
- [11] L. Rao, X. Liu, M. D. Ilic, and J. Liu, "Distributed coordination of Internet data centers under multi-regional electricity markets," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 269–282, Jan. 2012.
- [12] L. Wang, J. Zhan, W. Shi, and Y. Liang, "In cloud, can scientific communities benefit from the economies of scale?," *IEEE Trans. Parallel and Distributed Systems*, vol. 23, no. 2, pp. 296–303, Feb. 2012.
- [13] M. Parvania, M. Fotuhi-Firuzabad, and M. Shahidehpour, "Demand response participation in wholesale energy markets," in *Proc. 2012 IEEE PES General Meeting*, San Diego, USA, Jul. 2012.
- [14] M. Parvania and M. Fotuhi-Firuzabad, "Demand response scheduling by stochastic SCUC," *IEEE Trans. Smart Grid*, vol. 1, no. 1, pp. 89–98, Jun. 2010.
- [15] CAISO Day-Ahead Market LBMP [Online]. Available: <http://www.caiso.com>
- [16] NYISO Day-Ahead Market LBMP [Online]. Available: <http://www.nyiso.com>
- [17] ERCOT Day-Ahead Market LBMP [Online]. Available: <http://www.ercot.com>
- [18] A.-H. Mohsenian-Rad and A. Leon-Garci, "Energy-information transmission tradeoff in green cloud computing," in *Proc. IEEE GlobeCom '10*, Mar. 2010.
- [19] U.S. Energy Information Administration, Average Retail Price of Electricity to Ultimate Customers by End-Use Sector by State Mar. 2010.
- [20] US Environmental Protection Agency eGRID2012 Version 1.0 April 2012.
- [21] [Online]. Available: <http://www.google.com/about/datacenters/gallery/#/locations>
- [22] N. Gröwe-Kuska, H. Heitsch, and W. Römis, "Scenario reduction and scenario tree construction for power management problems," in *Proceedings of the IEEE Power Tech Conference*, Bologna, June 2003, vol. 3, pp. 23–26.
- [23] H. Heitsch and W. Römis, "Scenario reduction algorithms in stochastic programming," *Computational Optimization and Applications*, vol. 24, no. 2–3, pp. 187–206, Feb. 2003.
- [24] D. L. Keefer, "Certainty equivalents for three-point discrete-distribution approximations," *Management Science*, vol. 40, no. 6, pp. 760–773, Jun. 1994.
- [25] G. C. Pflug, "Scenario tree generation for multiperiod financial optimization by optimal discretization," *Mathematical Programming*, vol. 89, no. 2, pp. 251–271, Jan. 2001.
- [26] M. Kaut and S. W. Wallace, "Evaluation of scenario-generation methods for stochastic programming," *Stochastic Programming*, vol. 3, no. 2, pp. 257–271, May 2007.
- [27] J. Valenzuela and M. Mazumdar, "Monte Carlo computation of power generation production costs under operating constraints," *IEEE Trans. Power Syst.*, vol. 16, no. 4, pp. 671–677, Nov. 2001.
- [28] V. Sivaraman, A. Vishwanath, Z. Zhao, and C. Russell, "Profiling per-packet and per-byte energy consumption in the NetFPGA Gigabit router," in *Computer Communications Workshops (INFOCOM WKSHPS)*, IEEE, 2011, pp. 331–336.
- [29] Z. Chen, L. Wu, and Y. Fu, "Real-time price-based demand response management for residential appliances via stochastic optimization and robust optimization," *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 1822–1831, Dec. 2012.
- [30] R. Ayre, K. Hinton, and R. S. Tucker, "Energy consumption in wired and wireless access networks," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 70–77, Jun. 2011.
- [31] [Online]. Available: <http://www.google.com>
- [32] Considering a U.S. Carbon Tax: Frequently Asked Questions Resources for the Future, Dec. 2012 [Online]. Available: http://www.rff.org/centers/climate_and_electricity_policy/Documents/carbon-tax-FAQs.pdf
- [33] S. Benjaafar, Y. Li, and M. Daskin, "Carbon footprint and the management of supply chains: Insights from simple models," *IEEE Trans. Autom. Sci. Eng.*, vol. 10, no. 1, pp. 99–106, Jan. 2013.
- [34] S. Braithwait and K. Eakin, The Role of Demand Response in Electric Power Market Design, Laurits R. Christensen Associates. Madison, Oct. 2002 [Online]. Available: <http://www.eei.org>
- [35] E. Ela, B. Kirby, N. Navid, and J. C. Smith, "Effective ancillary services market designs on high wind power penetration systems," in *Proc. 2012 IEEE PES General Meeting*, San Diego, USA, Jul. 2012.
- [36] Climate Action Plan Tax City of Boulder-Colorado, Oct. 2009.
- [37] [Online]. Available: <http://www.gams.com>
- [38] L. Wu and M. Shahidehpour, "Accelerating benders decomposition for network-constrained unit commitment problems," *Energy Systems*, vol. 1, no. 3, pp. 339–376, Aug. 2010.

Zhi Chen (S'11) received the BE in Electronic Science and Technology from Wuhan University, China, in 2009, and MS degree in Electrical Engineering from Polytechnic Institute of New York University, USA, in 2011. He is currently working toward the Ph.D. degree at Clarkson University, USA. His research interests include demand response and the smart grid.

Lei Wu (M'07) received the BS in EE and MS in Systems Engineering from Xi'an Jiaotong University, China, in 2001 and 2004, respectively, and Ph.D. degree in EE from Illinois Institute of Technology, Chicago, USA, in 2008. From 2008 to 2010, he was a senior research associate in the Electric Power and Power Electronics Center at Illinois Institute of Technology. Presently, he is an Assistant Professor in the ECE Department at Clarkson University. His research interests include power systems operation and economics.

Zuyi Li (SM'09) received the BS in EE from Shanghai Jiaotong University in 1995, the MS in EE from Tsinghua University, China, in 1998, and the Ph.D. degree in EE from Illinois Institute of Technology, Chicago, in 2002. Presently, he is an Associate Professor in the Electrical and Computer Engineering Department at the Illinois Institute of Technology. His research interests include electricity market design and operation, renewable energy integration, and microgrid design and operation. Prof. Li is the Associate Director of the Robert W. Galvin Center for Electricity Innovation at IIT.