

Received November 30, 2021, accepted December 21, 2021, date of publication December 31, 2021, date of current version January 18, 2022.

Digital Object Identifier 10.1109/ACCESS.2021.3139985

Auction-Based Scheduling of Excess Energy Consumption to Enhance Grid Upward Flexibility

AHMED ABADA^{ID}, MARC ST-HILAIRE^{ID}, (Senior Member, IEEE),
AND WEI SHI^{ID}, (Member, IEEE)

Department of Systems and Computer Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada

Corresponding author: Ahmed Abada (ahmedabada@email.carleton.ca)

ABSTRACT Grid balancing is an essential requirement for power grid systems. This requirement has traditionally been fulfilled by existing flexibility mechanisms that provide voltage and frequency regulation. However, the recent interests in greening the energy supply by using more renewable energy sources present new grid balancing challenges. Such volatile energy sources introduce generation-side uncertainty and cause the flexibility mechanisms to fall short more often on providing enough balancing capacity. In this paper, we target the problem of balancing surplus energy from renewable sources by selling it in an auction to allow for its quick consumption. Our solution uses cloud datacenters as managed loads by incentivizing inter-datacenter cloud workload migrations through the auction sale of excess energy. It leverages the programmability and energy-demands flexibility of cloud datacenters, and uses incentivized cloud workload migrations to increase the energy consumption of a datacenter in a certain location to consume its excess energy. We propose an integrated auctioning-scheduling mechanism that auctions the surplus energy and schedules its consumption on a cloud datacenter. The auction part incentivizes the inter-datacenter cloud workload migrations, while the scheduling part ensures that migrated workloads do not exceed the destination datacenter capacity. Existing datacenter-based grid balancing approaches have focused on providing downward flexibility and only considered the case of owner-operated datacenters. In contrast, our system focuses on providing upward flexibility to target the increasingly frequent problem of excess energy from renewables, and it uses public cloud datacenters to increase the participation of datacenters in providing demand-side flexibility. Conducted simulation experiments show the effectiveness of our approach in ramping up the energy consumption of a target datacenter to minimize the time needed to consume the excess energy by as much as 75%. Plus, selling excess energy in an auction was shown to salvage 55%-65% of its original cost and provide 10% of cost savings to buyers.

INDEX TERMS Datacenters, energy auction, grid balancing, renewable energy sources (RESs), Vickrey-Clarke-Groves (VCG) mechanism, virtual machine (VM) migration.

I. INTRODUCTION

Grid-balancing is essential for the proper operation of power grids. Since they lack large-scale energy storage capabilities [1], [2], their supply and demand levels must always be kept equal in order to avoid system instability, power interruptions, and frequency deviations that are harmful to end-users' equipment. Therefore, power grids must always have enough flexibility resources that allow them to adjust the energy levels on the generation and consumption sides in response to possible imbalance events [3], [4]. For grid

balancing, the term "Upward Flexibility" refers to a system's ability to increase its generation in response to increasing demand or increase its demand in response to increasing generation. Similarly, "Downward Flexibility" refers to a system's ability to decrease its generation in response to slowing demand or decrease its demand in response to decreasing generation. When an imbalance event materializes, the ability of a system to stay balanced depends on the availability of the needed flexibility resources.

A. MOTIVATION AND BACKGROUND

Traditionally, power grid operators maintained grid balance by producing just enough energy to meet expected

The associate editor coordinating the review of this manuscript and approving it for publication was Behnam Mohammadi-Ivatloo^{ID}.

demand and used generation-side flexibility resources to adjust against real-time demand variations. First, demand prediction approaches are used to forecast the demand of electricity customers over future time intervals [5], [6]. Then, Generation-units scheduling approaches like Unit-Commitment (UC) and Economic Dispatch (ED) are used to optimize the number of active generation-units and their operational set-points to match their produced energy to expected demand [7], [8]. After fixing the generation-units' output, real-time demand variations are then balanced using marginal amounts of generation-side flexibility resources. Such flexibility resources are usually in the form of quick-response backup generation capacity running under the Automatic Generation Control (AGC) protocol that adjusts their output in response to changes in demand to provide real-time upward/downward flexibility [9]. More recently, new smart-grid advancements have allowed utility companies to introduce Demand Response (DR) programs to achieve grid balance by exploiting demand-side flexibility [10], [11]. DR programs refer to a wide range of methods that use monetary incentives to influence energy demand modifications of large-scale consumers.

The above combination of grid-balancing measures (prediction-based energy generation scheduling, flexibility resources, and DR) has effectively maintained grid balance for systems using conventional power generation. However, the recent trends of greening the power supply by integrating more Renewable Energy Sources (RESs) into the power grid introduce generation side uncertainty and necessitate increasing the flexibility resources' capacity to ensure continued reliable operation. Figure 1 illustrates the effect of increasing the adoption of solar energy in recent years on a typical 24-hour net-demand curve (the net amount of energy produced by conventional sources during a 24-hour horizon after taking into account renewable generation input) as seen by the California Independent System Operator (CAISO) [12]. It shows that increasing the amount of solar generation reduces the load on conventional generators during the daylight hours (between 7:00 AM and 5:00 PM). However, as the electricity demand increases in the evening hours for cooking, watching TV, Etc., the power grid would need to quickly ramp up its generation from conventional sources to meet the increasing demand while solar generation wanes. The increased flexibility requirements shown above can be met in multiple ways. One option is to use more generation-side flexibility resources. Another alternative is to upgrade the transmission network to enhance its ability to distribute load among multiple areas and reduce the effect of generation fluctuations [13], [14]. Lastly, the use of DR can be expanded to gain more demand-side flexibility. The first option of increasing the amount of generation-side flexibility resources is not practical because of its associated high cost and the large amount of additional capacity that would need to be allocated to achieve the needed flexibility [15]–[17]. The second option is also difficult to implement because such infrastructure upgrades are very costly. The last option of

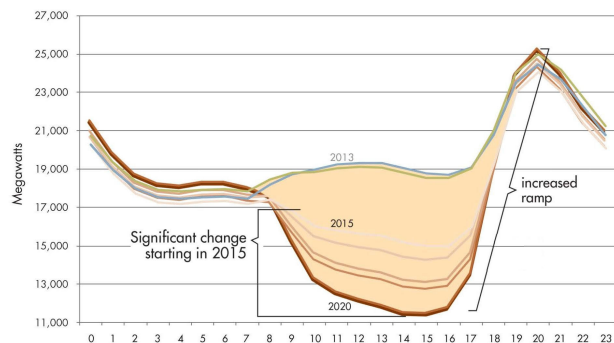


FIGURE 1. Net load curve after RESs integration [12].

expanding DR capacity is the most viable for enhancing the overall system flexibility because it is only used as needed and is more economical compared to the other alternatives.

Given the need for increased demand-side flexibility, many DR proposals were introduced to allow using different kinds of industrial consumers as managed loads [18], [19]. Among the new DR proposals, many have focused on developing new approaches that target datacenters. Datacenters consume large amounts of energy [20], [21] and have elastic energy demands since their energy consumption is always proportional to the amount of workload they handle [22]–[24]. Such properties make them suitable candidates for DR to provide the needed demand-side flexibility. However, existing datacenter-based DR research did not address two important aspects: First, they only considered the case of “Owner-operated” datacenters where datacenter operators control the execution and scheduling of computational workloads. This assumption does not apply to a growing number of Infrastructure as a Service (IaaS) public cloud datacenters where the scheduling of computational workloads is controlled by end-users and cloud-brokers that issue the tasks [25]. Cloud brokers are large-scale cloud aggregators/resellers that offer added-value services to cloud end-users and charge a markup to generate profit. They do not own datacenter infrastructure resources for hosting end-users' requests. Hence, they rent physical resources from IaaS cloud datacenters to host the Virtual Machines (VMs) needed to process their clients' workloads. Second, existing research has only focused on the traditionally common case of downward-flexibility that tries to minimize consumers' energy demand at times of generation shortages. However, practical experience has shown that increasing the deployment of RESs into power grids has made excess energy generation events more frequent, especially at higher shares of RESs' deployment [26], [27], leading to energy market instabilities and negative energy prices in real-time balancing markets [28]. Therefore, enhancing the grid's upward flexibility is necessary to balance excess RESs' energy.

As illustrated by the above research gaps, our work is motivated by two main challenges. The first is to introduce a new upward flexibility mechanism to address the rising frequency of excess energy imbalance events. The second

is to devise a new DR approach suitable for public IaaS cloud datacenters to allow them to participate in providing demand-side flexibility. We address the first challenge by introducing a new market mechanism that sells excess energy in an auction to incentivize its quick consumption. We address the second challenge by enabling the participation of cloud brokers in the proposed system to control the energy demand of IaaS cloud datacenters through workload migrations.

Recent research has emphasized the importance of developing new energy market mechanisms that enable the participation of end-users in DR programs [29], [30]. Enabling end-users' participation in DR allows for introducing new DR designs and is especially necessary in scenarios when energy-consumption decisions are made by end-users, as in the case of IaaS public cloud datacenters. To that end, we introduce a new energy market mechanism that sells excess energy in an auction to cloud brokers. The purpose of the auction is to incentivize brokers to migrate their cloud workloads to a destination IaaS cloud datacenter in a location of excess energy to increase its energy consumption and consume the excess energy.

B. CONTRIBUTIONS

The main contributions and novelties of our work are the following:

- We propose a new auction-based market mechanism for selling excess RESs' energy to motivate its quick consumption and provide grid upward flexibility.
- We use public IaaS cloud datacenters to provide grid upward flexibility through cloud workload migrations. To the best of our knowledge, this is the first work to use public IaaS cloud datacenters for providing grid upward flexibility. Others have only considered owner-operated datacenters.
- We develop a combined auctioning-scheduling optimization model that jointly auctions the excess energy and schedules its consumption on a destination datacenter. Most other models do this separately.

Unlike existing methods for balancing excess RESs' energy that either use expensive generation-side flexibility resources [3], [4], [31], require transmission network upgrades [13], [14], or sell the excess energy at negative prices in the real-time balancing market [28], our solution is more cost efficient since it does not require allocating additional flexibility resources that may not always be used, does not require making huge transmission network upgrades or lose money by selling excess energy at negative prices. The proposed system is only used as needed when an excess energy imbalance event materializes. It generates money from the auction sale as opposed to losing money by selling excess energy at negative prices. Compared to using other types of managed loads for providing demand-side flexibility [18], [19], using datacenters workload migrations has the effect of transferring energy load across geographical areas as the energy consumption is transferred from one datacenter location to another when workload is migrated. Therefore,

bypassing possible transmission network bottlenecks and helping delay the need for transmission network upgrades.

C. PAPER ORGANIZATION

This paper is organized as follows: In Section II, we summarize the main previous works in the area of datacenter-based DR research and we compare it to our work. In Section III, we introduce the system model, the used cloud workloads pricing models and the system's objective. In Section IV, we present the proposed auctioning-scheduling mechanism in detail, followed by its performance evaluation in Section V. We then discuss the limitations of the proposed approach in Section VI and we conclude in Section VII.

II. LITERATURE REVIEW

There has been much research on the cooperation between datacenters and the grid using DR programs for the purpose of grid balancing. However, they only considered the case of downward flexibility (to reduce energy consumption and avoid the need for additional generation capacity at times of high-demand) and used owner-operated datacenters. We review in this section the current related-work and illustrate in Table 1 the novelty of our approach compared to existing works. Ghamkhari *et al.* [32] rely on using local energy storage to take advantage of varying energy prices. They propose forming DR contracts between datacenters and their supplying utility grids, by which, a datacenter must decrease its energy consumption level upon receiving demand signals from the grid. Chen *et al.* [33] investigate the feasibility of using datacenters to offer DR via server consolidation and dynamic server power capping. Nguyen *et al.* [34] design a non-cooperative game framework for minimizing the maximum Peak-to-Average (PTA) energy drawn from a grid. They consider a set of grid clients equipped with on-site batteries, each with a different energy demand over time. They use price signals (sent by the grid) to control the energy consumption for minimizing the maximum PTA ratio, while the grid clients on the other hand try to minimize their total energy cost. Liu *et al.* [35] consider a grid system with renewable energy sources and try to optimize the incentives given to datacenters for their participation in DR. They use an online prediction-based method to generate estimates for future energy availability and user demands. Cao *et al.* [36] assumes a non-cooperative game scenario between a power grid and a datacenter. They develop a game theory bargaining approach for settling the reward price for load reduction by the datacenter. Ma *et al.* [37] consider the problem of selecting a minimum number of energy customers (from a given pool) to provide a required amount of demand reduction with a certain level of reliability. They propose two reward-based mechanisms that illicit from the customers either the amounts of energy reductions they are capable of and their sought payments or their willingness to accept penalties if they do not deliver on their promised reductions. Their approach tries to provide reliable delivery of energy reduction and minimize the resulting cost (payments to the selected customers).

TABLE 1. Summary of related literature and current work.

Research	Model parameters	Datacenter type	Flexibility direction
[32]	- Local energy storage - Variable energy prices	Owner operated	downward
[33]	- Server consolidation - Dynamic server power capping	Owner operated	downward
[34]	- Price signals - On-site batteries	Owner operated	downward
[35]	- Variable energy prices - Future workload prediction	Owner operated	downward
[36]	- Reward for datacenter load reduction	Owner operated	downward
[37]	- Reward for datacenter load reduction - Ensures reliable load reduction	Owner operated	downward
[38]	- Dynamic energy pricing at different locations to control demand	Owner operated	downward
Current work	- Sale of excess energy in an auction - Uses cloud workload migrations	Public IaaS	upward

Wang *et al.* [38] study the grid-datacenter interaction and propose a two stage model that targets both grid balancing and minimizing the datacenter energy cost. In their model, the grid uses dynamic electricity pricing to control the energy demand of datacenters in order to balance the grid, while the datacenters try to optimize their workload scheduling to minimize their energy cost according to the variable energy prices. They formulate the problem as a linear program and propose a branch-and-bound algorithm to attain a globally optimal solution.

III. SYSTEM MODEL

This section presents the proposed energy auctioning-scheduling system that can be used to increase the energy consumption of a datacenter within a grid balancing zone to consume its excess RESs' energy. To the best of the authors' knowledge, this work is the first to use datacenters to provide grid upward-flexibility via inter-datacenter cloud workload migrations. Since the main entities considered in this work are the cloud brokers and datacenters, we modeled our system as a set I of cloud brokers and a set J of datacenters (the IaaS cloud providers). Each datacenter is located within a different balancing zone, as shown in Figure 2. Each balancing zone is managed by a local utility company (the power grid in Figure 2). Utility companies use conventional and RESs generation units to produce energy. They are also responsible for maintaining grid balance within their balancing zones using the methods introduced in Section I (prediction-based energy generation scheduling, flexibility resources, and DR).

We focus on the case of when RESs' output is inaccurately predicted, leading to an imbalance event where excess energy must be consumed to maintain grid balance. In such a case, the local utility would auction the excess energy to the cloud brokers (as shown in the next section) to incentivize its quick consumption. We assume that cloud brokers pay for the energy consumption of their workloads on IaaS datacenters (as shown in Subsection III-B); therefore, they are motivated to migrate their workloads to a datacenter offering a cheaper energy price to lower their cost. Since excess energy is not sold at its regular price in the auction, it is represented as

Energy Credits (ECs). ECs are energy coupons that can be used to pay for energy usage at the datacenter in the issuing utility's balancing zone.

To streamline the ECs' auctioning process, we introduced a centralized ECs' market (shown in Figure 2) to auction the ECs for the local utilities. The ECs' market is connected to the other system components (brokers, datacenters, and local utilities) to facilitate the auctioning task. When a local utility wants to consume a certain amount of excess energy within its balancing zone, it initiates a new ECs' auction by sending the ECs it wants to auction to the ECs' market. The ECs' market would then announce a new auction to all brokers, collect their bids, run the auction on behalf of the utility that issued the ECs and determine the allocation result. ECs are sold in an auction with no reserve price in order to receive bids from all brokers with positive valuations and minimize the time needed to consume the excess energy. In addition, the auction requires that the sold ECs may only be used at the intended datacenter within the balancing zone of excess energy, thereby incentivizing cloud brokers to migrate their workloads to that datacenter to take advantage of auctioned energy pricing.

The current centralized design simplifies the joining process for new utilities and datacenters wishing to join the system as they would only need to connect to the ECs' market to join. It also simplifies the discovery process for brokers wishing to learn about new ECs' auctions since auction announcements are always broadcasted to all participating brokers by the centralized ECs' market. Lastly, the system-wide view of the centralized ECs' market allows it to compute globally optimal allocations.

We assume for simplicity that a maximum of one ECs' auction may take place in the system at any given time. Therefore, there can be a maximum of one balancing zone with excess energy at any given time. This assumption was made to simplify the ECs' market operation and focus on the main idea of using auctions to provide upward flexibility via cloud datacenters workload migrations. The complete list of variables used in our model is presented in Table 2.

A. BROKERS PRICING MODEL FOR END-USERS WORKLOADS

Cloud brokers receive workload requests from their clients (the end-users) on a random arrival basis, where each request is described by a four-attribute tuple $W = (w^c, w^m, w^s, w^d)$ as the computing, memory, storage and duration requirements of the submitted workload. Brokers charge the end-users for their submitted workload requests based on the cost parameters $RateB_i^c$, $RateB_i^m$ and $RateB_i^s$ (that represent the per-unit cost of processing, memory and storage resources respectively) for the required execution duration. Therefore, the total cost charged by a broker i to an end-user u for processing a workload W_{ui} can be calculated as the following:

$$CostWB_{ui} = (RateB_i^c * w_{ui}^c + RateB_i^m * w_{ui}^m + RateB_i^s * w_{ui}^s) * w_{ui}^d \quad (1)$$

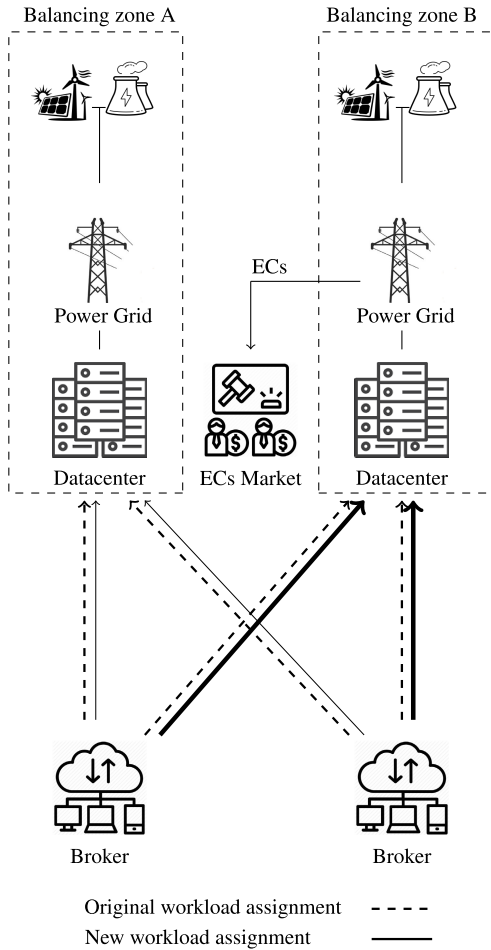


FIGURE 2. Energy Credits Market Model.

B. DATACENTERS PRICING MODEL FOR BROKERS WORKLOADS

Since the cloud brokers act as aggregators for end-users demand, we assume they are given a special pricing plan (for high-volume customers) at the datacenters. Besides allowing the brokers to generate profit, such a pricing plan also recognizes the important role they can play in making the datacenters’ energy consumption more responsive to DR signals by exposing them to the electricity cost component that is usually hidden from smaller consumers for simplicity. This incentivizes the cloud brokers to participate in DR programs to minimize their costs and help maintain grid balance. A similar approach was attempted by Wang *et al.* in [39], albeit with different implementation details. Under this cloud brokers’ pricing model, datacenters apply the following cost parameters $RateD^c$, $RateD^m$, $RateD^s$ and $RateD^e$ as their per-unit costs of computing, memory, storage and energy resources consumed by the brokers’ workloads respectively. Since the energy consumption of VMs depends on the type of physical hardware used, cloud brokers must calculate the anticipated energy consumption of their workloads on each datacenter before making workload placement/migration decisions. To make this calculation possible, datacenters

provide the brokers with their hardware-specific energy consumption parameters $ConvD^c$, $ConvD^m$, $ConvD^s$ that represent the amount of energy consumed per unit of processing, memory, and storage respectively allocated per time interval, which is a widely accepted energy consumption model for computing hardware [40]. Therefore, the total cost charged by a datacenter j to a broker i for processing a workload W_{ij} can be calculated as the following:

$$\begin{aligned}
 CostWD_{ij} &= \left[RateD_j^c * w_{ij}^c + RateD_j^m * w_{ij}^m \right. \\
 &\quad \left. + RateD_j^s * w_{ij}^s + RateD_j^e * (ConvD_j^c * w_{ij}^c \right. \\
 &\quad \left. + ConvD_j^m * w_{ij}^m + ConvD_j^s * w_{ij}^s) \right] * w_{ij}^d \quad (2)
 \end{aligned}$$

While current public IaaS datacenters do not disclose the energy consumption parameters of their equipment as shown in the above pricing model, there is a good business case for them to do so under the proposed system to enter into a mutually beneficial collaboration with the brokers. Taking part in this system would mean more business for the datacenters at times of energy surplus since they can potentially increase their sales by hosting migrated workloads. Brokers also benefit by participating in the proposed system since they always lower their costs when migrating their workloads. Additionally, the hardware-specific energy consumption parameters are scalar values related to the energy ratings of computing equipment and are not part of program data. Therefore, the potential for compromised security to end-users or datacenters due to sharing the values of such parameters with the brokers is very low.

C. SYSTEM OBJECTIVE

The system’s main objective is to consume/balance a certain amount of excess RESs’ energy by creating an equal demand for it on the consumption side using inter-datacenter cloud brokers’ workload migrations. To create this demand, we incentivize brokers to migrate a portion of their workloads to a datacenter within the grid balancing zone that has the excess RESs’ energy by selling the excess energy in an auction. The desired behavior of the system is to maintain a real-time balance between supply and demand during RESs’ oversupply events.

In order to focus on the main idea of the proposed approach, we assume ideal workload migration conditions where the workloads’ migration time and cost are negligible. This simplifies the system model and is not a huge departure from practical scenarios since current advancements in VM migration technologies enable negligible-cost migration and allow it to happen seamlessly in the background without affecting end-users’ Quality of Service (QoS) [41]. Nevertheless, the proposed system can still accommodate any migration-related costs by simply adding such costs into Eq. 4 for calculating $CostD_{ijk}$.

TABLE 2. The scheduling auction decision variables.

Symbol	Meaning
Sets and Indices	
I	Set of brokers where $i \in I$
J	Set of datacenters where $j, k \in J$
U	Set of end-users where $u \in U$
Broker Variables	
$RateB_i^c$	Processing cost of broker i
$RateB_i^m$	Memory cost of broker i
$RateB_i^s$	Storage cost of broker i
Bid_{ijk}	Tuple of the form: $Bid_{ijk} = (w_{ijk}^c, w_{ijk}^m, w_{ijk}^s, w_{ijk}^e, v_{ijk})$ that represents the bid of broker i to migrate its workload from datacenter j to datacenter k . It specifies the amounts of resources needed at the destination datacenter (processing (w_{ijk}^c), memory (w_{ijk}^m), storage (w_{ijk}^s) and energy (w_{ijk}^e)) and how much it is willing to pay for them per time interval (v_{ijk})
Datacenter Variables	
$Capacity_{kh}^c$	Max CPU capacity at datacenter k during interval h
$Capacity_{kh}^m$	Max memory capacity at datacenter k during interval h
$Capacity_{kh}^s$	Max storage capacity at datacenter k during interval h
$RateD_k^c$	Processing cost at datacenter k
$RateD_k^m$	Memory cost at datacenter k
$RateD_k^s$	Storage cost at datacenter k
$RateD_k^e$	Energy cost at datacenter k
$ConvD_k^c$	Energy consumed by one unit of processing resources allocated at datacenter k per time interval
$ConvD_k^m$	Energy consumed by one unit of memory resources allocated at datacenter k per time interval
$ConvD_k^s$	Energy consumed by one unit of storage resources allocated at datacenter k per time interval
EC_k	Number of energy credits at datacenter k
Workload Variables	
W_{ij}	Total broker i workloads handled by datacenter j , $W_{ij} = (w_{ij}^c, w_{ij}^m, w_{ij}^s, w_{ij}^d)$
$w_{ij}^c, w_{ij}^m, w_{ij}^s$	CPU, memory and storage components of W_{ij} respectively
w_{ij}^d	The duration of W_{ij}
w_{ij}^e	Energy consumed by broker i workloads on datacenter j per time interval
Decision Variables	
A_{ijkh}	Binary matrix that represents the allocated execution intervals for the accepted received bids
S_{ijkh}	Binary matrix that represents the starting time of executing the accepted received bids
P_{ijk}	The payment amount charged to a winning bid Bid_{ijk} as per the VCG payment rule

IV. MULTI-INTERVAL AUCTION MECHANISM

Since the auction sale of ECs must consider the scheduling of migrated cloud workloads, we use an iterative lookahead combined auctioning-scheduling formulation with a lookahead size of T intervals. The auctioning-scheduling optimization is iteratively executed at the beginning of each interval to auction the available ECs and update the workloads scheduling allocation for the next T intervals. We chose to implement the famous Vickrey-Clarke-Groves (VCG) auction mechanism in our model since it is a Dominant Strategy Incentive Compatible (DSIC) mechanism by definition, proven to provide strong incentive compatibility guarantees (i.e., guarantees that rational bidders will always bid their true valuations) [42]. This greatly simplifies the brokers' bids generation process, makes it easy to automate, and guarantees positive utility on accepted bids.

The proposed ECs' auction mechanism introduced here is carried out when a certain amount of excess RESs' energy needs to be consumed at a certain balancing zone to maintain its energy balance. The auction mechanism manages the sale of excess RESs' energy and the scheduling of migrated workloads on the available datacenter resources. Selling the excess RESs' energy in an auction allows the grid to consume the oversupply quickly while generating revenue (as opposed to simply discarding the excess energy). The process starts when a certain grid balancing zone offers its excess RESs' energy (represented as ECs) for sale in an auction at the ECs' market (as shown in Figure 2). The ECs' market then invites brokers to submit bids for buying fractions of the available ECs, receives bids from interested brokers, determines the winning bids and their payments. The winning brokers can then use the won ECs to pay for energy costs (i.e., to compensate against the term multiplied by $RateD_j^e$ in Eq. 2) at the target destination datacenter when their workloads are migrated there.

A. BIDS GENERATION

We assume that cloud brokers normally choose the minimum cost datacenter for hosting their clients' workloads, and they have workloads of different durations already hosted at different datacenters before the auction process begins. A new auction round starts when new ECs are introduced in the system by a certain balancing zone. Brokers are then invited to place bids for buying the amounts of energy that their migratable workloads would consume at the destination datacenter where the ECs can be used. Those amounts are easily computed for the migratable workloads using the publicly available hardware-specific energy consumption parameters ($ConvD_k^c$, $ConvD_k^m$ and $ConvD_k^s$) at the desired destination datacenter k . We refer to a datacenter from where a broker's workload is to be migrated from as the "source datacenter", and the datacenter to where the workload is to be migrated to as the "destination datacenter". It is important to note that in order for a broker to gain positive utility (i.e., not lose money) on any accepted bid, the maximum total cost of running its migratable workloads at a destination datacenter can not exceed their current cost at the source datacenter (as shown in Figure 3). We let $W_{ij} = (w_{ij}^c, w_{ij}^m, w_{ij}^s, w_{ij}^d)$ be a tuple that describes the amount of resources (CPU, memory and storage) and the duration of the workloads of broker i currently hosted at a datacenter j . Brokers generate their bids in the following tuple format: $Bid_{ijk} = (w_{ijk}^c, w_{ijk}^m, w_{ijk}^s, w_{ijk}^d, w_{ijk}^e, v_{ijk})$, which describes a bid submitted by a broker i for migrating an aggregated amount of clients workloads (of the same execution time requirements) from a source datacenter j to a destination datacenter k . In the above tuple, w_{ijk}^c , w_{ijk}^m , w_{ijk}^s are the amounts of CPU, memory and storage resources respectively that would be migrated from a source datacenter j to a destination datacenter k if Bid_{ijk} is accepted. w_{ijk}^d represents the workload duration. w_{ijk}^e represents the amount of energy that would be consumed per time interval at the destination datacenter k as a result of the migration (computed using the hardware-specific energy consumption parameters

as described earlier). This ensures that the amount of energy submitted in a bid is exactly the amount that would be consumed if the bid is accepted and the migration takes place. Lastly, since the other cost components of each datacenter ($RateD^c$, $RateD^m$, $RateD^s$) shown in Eq. 2 are also publicly available, the brokers can compute the value term (v_{ijk}) in their bids as the maximum price they are willing to pay for w_{ijk}^e at the destination datacenter k . To calculate how much a broker i would be willing to bid for a certain amount of energy w_{ijk}^e at a destination datacenter k while ensuring to maintain positive utility, we first compute the cost of running W_{ij} at the destination datacenter k without including the energy cost portion (assuming that broker i would not have to pay for energy at the destination datacenter k in order to get a lower bound on cost there). Then, we subtract this value from the cost of what broker i is currently paying for W_{ij} at the source datacenter j . This gives us the maximum positive utility that broker i could generate by migrating W_{ij} to the destination datacenter k if its bid Bid_{ijk} is accepted. From Eq. 2, the current cost for broker i for running its workload W_{ij} at a source datacenter j can be computed as the following:

$$\begin{aligned}
 CostS_{ijk} &= RateD_j^c * w_{ijk}^c + RateD_j^m * w_{ijk}^m \\
 &\quad + RateD_j^s * w_{ijk}^s + RateD_j^e * (ConvD_j^c * w_{ijk}^c \\
 &\quad + ConvD_j^m * w_{ijk}^m + ConvD_j^s * w_{ijk}^s) \quad (3)
 \end{aligned}$$

On the other hand, the cost for broker i for running its workload W_{ij} at a destination datacenter k without considering the energy cost there is the following:

$$\begin{aligned}
 CostD_{ijk} &= RateD_k^c * w_{ijk}^c + RateD_k^m * w_{ijk}^m \\
 &\quad + RateD_k^s * w_{ijk}^s \quad (4)
 \end{aligned}$$

Therefore, v_{ijk} can be calculated as shown in Figure 3 as the following:

$$v_{ijk} = CostS_{ijk} - CostD_{ijk} \quad (5)$$

The bids' generation process is locally executed by each broker and is summarized in Algorithm 1. We assume that bidders are single-minded, meaning that they do not accept partial allocations on their bids and derive zero utility from such allocations. Consequently, their valuation v_{ijk} on any partial allocation of w_{ijk}^e is zero.

B. WINNERS DETERMINATION/RESOURCE ALLOCATION

The VCG auction mechanism is a social surplus maximizing mechanism. Therefore, its Winners' Determination Problem (WDP) is solved by selecting the bids that maximize the total bidders' valuation while satisfying applicable problem-specific constraints. Since the proposed multi-interval scheduling auction is concerned with both of auctioning the available ECs and determining the execution schedule of the migrated brokers' workloads, we use the two decision variables S_{ijkh} and A_{ijkh} as our auction optimization result to represent the execution schedule of the accepted

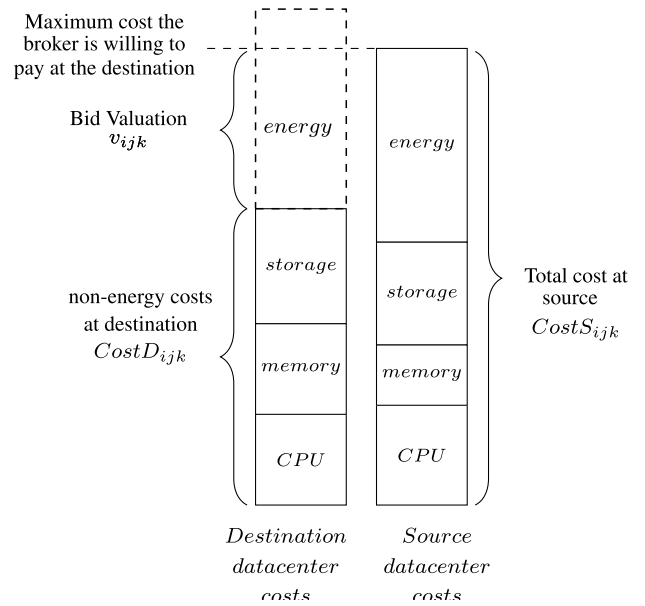


FIGURE 3. Bid Calculation.

bids during the lookahead time horizon T . Both variables are a four-dimensional boolean array. They indicate the auction mechanism decision on accepting/rejecting the bids submitted by the brokers for migrating workloads of a broker i from a datacenter j to a datacenter k , and the allocated execution intervals h .

The array S_{ijkh} encodes the starting interval (during the lookahead time horizon T) for executing the migratable workload W_{ijk} on datacenter k if Bid_{ijk} is accepted and the migration is to take place. On the other hand, the array A_{ijkh} encodes the allocated intervals for executing the workload (which must be uninterruptible with length equal to the required execution duration w_{ijk}^d). An example outcome of S and A for an accepted bid Bid_{ijk} of a broker i is shown in Figure 4, where a workload of duration $w_{ijk}^d = 3$ is accepted to be migrated from datacenter $j = 3$ to datacenter $k = 4$ and is scheduled to start execution in the time interval $h = 2$ for three consecutive intervals. It is noted that while either of the above arrays would suffice to deduce the full information about the auction output and the scheduling outcome, they are both needed to properly construct the integer linear programming constraints in our model that guarantee uninterruptible execution of migrated workloads. The WDP is solved by executing the following auctioning-scheduling combined optimization problem using the received bids Bid_{ijk} as input, where k is the index of the destination datacenter:

Objective function:

$$\text{Max} \sum_{i \in I} \sum_{j \in J} \sum_{k \in J} \sum_{h \in T} S_{ijkh} * v_{ijk} * w_{ijk}^d \quad (6)$$

$h < (|T| - w_{ijk}^d + 1)$

Subject to:

Capacity Constraints:

$$\sum_{i \in I} \sum_{j \in J} w_{ijk}^c * A_{ijkh} < Capacity_{kh}^c \quad (h \in T) \quad (7)$$

Algorithm 1 Algorithm for Bids Generation by Cloud Brokers (performed Individually by Each Cloud Broker $i \in I$ at the Beginning of Each Time Interval When ECs Are Available for Auction at Datacenter k)

Input: i : index of cloud broker
 j : index of source datacenter
 k : index of destination datacenter
 J : set of available datacenters
 W_{ij} : workload of cloud broker i on datacenter j , $W_{ij} = (w_{ij}^c, w_{ij}^m, w_{ij}^s, w_{ij}^d)$
Output: Bid_{ijk} : bid to migrate workload W_{ij} of broker i from datacenter j to datacenter k

Initialisation : broker i receives an announcement (from the ECs market) about ECs availability at datacenter k

- 1: **for** $j = 0$ to $|J|$ **do**
- 2: **if** (i has W_{ij} on datacenter j) **and** ($j \neq k$) **then**
- 3: compute $CostS_{ijk}$ (Eq. 3)
- 4: compute $CostD_{ijk}$ (Eq. 4)
- 5: compute v_{ijk} (Eq. 5)
- 6: compute $w_{ijk}^e = (ConvD_k^c * w_{ijk}^c + ConvD_k^m * w_{ijk}^m + ConvD_k^s * w_{ijk}^s)$
- 7: compute $w_{ijk}^d = (w_{ijk}^c, w_{ijk}^m, w_{ijk}^s, w_{ijk}^d, w_{ijk}^e, v_{ijk})$
- 8: send Bid_{ijk} to ECs' market
- 9: **end if**
- 10: **end for**
- 11: **end for**

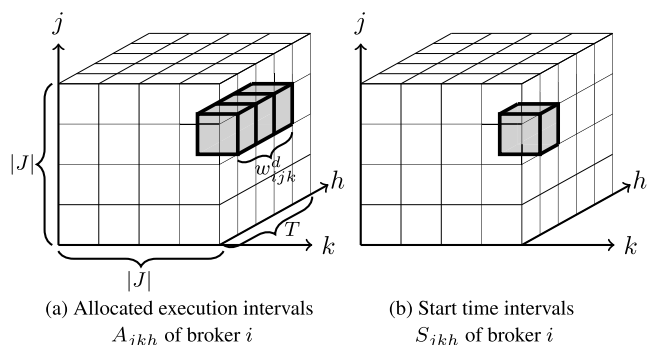


FIGURE 4. The scheduling auction decision variables.

$$\sum_{i \in I} \sum_{j \in J} w_{ijk}^m * A_{ijkh} < Capacity_{kh}^m \quad (h \in T) \quad (8)$$

$$\sum_{i \in I} \sum_{j \in J} w_{ijk}^s * A_{ijkh} < Capacity_{kh}^s \quad (h \in T) \quad (9)$$

$$\sum_{i \in I} \sum_{j \in J} \sum_{h < (|T| - w_{ijk}^d + 1)} S_{ijkh} * w_{ijk}^e < ECs \quad (10)$$

Continuity Constraints:

$$\sum_{i \in I} \sum_{j \in J} \sum_{h < (|T| - w_{ijk}^d + 1)} (-w_{ijk}^d * S_{ijkh}) + \sum_{i \in I} \sum_{j \in J} \sum_{h \in T} A_{ijkh} = 0 \quad (11)$$

$$\sum_{i \in I} \sum_{j \in J} \sum_{h < (|T| - w_{ijk}^d + 1)} \sum_{d=0}^{w_{ijk}^d} A_{ijk(h+d)} - \sum_{i \in I} \sum_{j \in J} \sum_{h < (|T| - w_{ijk}^d + 1)} (S_{ijkh} * w_{ijk}^d) \geq 0 \quad (12)$$

Single Allocation Constraint:

$$\sum_{h \in T} S_{ijkh} \leq 1, \quad \forall i \in I, j \in J \quad (13)$$

Total Allocation Constraint:

$$\sum_{h \in T} A_{ijkh} \leq w_{ijk}^d, \quad \forall i \in I, j \in J \quad (14)$$

Integrality Constraint:

$$S_{ijkh}, A_{ijkh} \in \{0, 1\}, \quad \forall i \in I, j \in J, k \in J, h \in T \quad (15)$$

The objective function (Eq. 6) in the above formulation is derived from the definition of the VCG auction, as it maximizes the total valuation on the accepted bids (bids are accepted if their corresponding $S_{ijkh} = 1$). The following capacity constraints (Eqs. 7-10) ensure that the total amount of resources' requirements in accepted bids do not exceed the available datacenter capacities and the amount of sold ECs is less than the available ECs. Continuity constraints (Eqs. 11,12) ensure that each accepted bid is allocated a number of consecutive time intervals equal to its duration. The single allocation constraint (Eq. 13) ensures that each bid can only be accepted at most once. The total allocation constraint (Eq. 14) ensures that each bid is not allocated a number of intervals greater than its duration. Finally, the integrality constraint (Eq. 15) ensures that no partial allocations are allowed on submitted bids by restricting the decision variables to the integer values of 0 and 1. The continuity and total allocation constraints work together to ensure that all accepted bids are allocated contiguous blocks of time intervals with lengths equal to their required execution durations w_{ijk}^d . The above formulation for winner-determination and resource allocation is summarized in Algorithm 2, which is executed at the beginning of each time interval by the ECs' market to process incoming bids Bid_{ijk} when ECs are auctioned.

C. PAYMENT CALCULATIONS

The payment rule of the VCG auction mechanism charges each winning bid Bid_{ijk} a payment p_{ijk} equal to its social cost. The social cost of a bidder i is defined as the maximum total utility that other bidders would have been able to receive if Bid_{ijk} did not participate minus their maximum utility when it participates. This payment calculation method requires solving a new optimization sub-problem (under the same original constraints of Eqs. 7-15) to calculate the payment for each winning bidder (to determine the maximum utility of the other bidders when i does not participate). The calculation

Algorithm 2 Algorithm for Auctioning the Available ECs and Scheduling Execution of Migrated Workloads on a Destination Datacenter k . (Performed by the ECs Market at the Beginning of Each Time Interval)

Input: EC : auctioned energy credits at datacenter k
 Bid_{ijk} : bid of cloud broker i to migrate its workload W_{ij} from datacenter j to datacenter k
 $Capacity_{kh}^{c,s,m}$: CPU, memory and storage capacity of destination datacenter k at time interval h

Output: S_{ijkh} : boolean decision variable, the starting time interval for migrated workload W_{ijk}
 A_{ijkh} : boolean decision variable, the allocated time interval for migrated workload W_{ijk}

- 1: receive bids Bid_{ijk} from brokers (as computed in Algorithm 1)
- 2: find optimal solution for (6) s.t. (7 - 15)
- 3: **return** S_{ijkh}, A_{ijkh}

of a payment p_{ijk} for a winning bidder on its bid Bid_{ijk} can be computed according to Eq. 16 as the following:

$$p_{ijk} = \max \sum_{i \in I} \sum_{j \in J} (S_{ijkh} * v_{(ijk)_{-1}} * w_{ijk}^d) - \left[\left(\max \sum_{i \in I} \sum_{j \in J} S_{ijkh} * v_{ijk} * w_{ijk}^d \right) - v_{ijk} \right] \quad (16)$$

In the above equation, the first term represents the maximum total valuation of winning bids when Bid_{ijk} is not taken into consideration, where $v_{(ijk)_{-1}}$ refers to the valuations of bids other than Bid_{ijk} . The second term calculates the maximum valuations of winning bidders other than Bid_{ijk} when Bid_{ijk} is taken into consideration. This is simply calculated by subtracting v_{ijk} from the optimal solution computed for the objective function during the winners' determination step. The above payments calculation process is executed by the ECs' market and is summarized in Algorithm 3.

D. COMPUTATIONAL COMPLEXITY

The VCG mechanism is currently the only known technique for designing truthful incentive-compatible auctions [42], [43]. It was chosen to implement the auction part in this work to simplify the brokers' bids-generation process because it guarantees that truthful bidding is a dominant strategy for all bidders wishing to maximize their utility. Therefore, eliminating the need for a bidding strategy. However, the truthfulness property of the VCG mechanism is only maintained if its WDP is solved to optimality [44], which is an NP-hard problem in the case of the proposed auction model.

Since we do not allow partial allocations on submitted bids (as bidders have zero valuation for partial allocations), the possible values for the decision variables are restricted to the integer values of 0 and 1 (Eq. 15). This makes the problem non-convex by definition [45]. Therefore, convex optimization approaches are not applicable and convex relaxations would not help since they would only yield

Algorithm 3 Algorithm for Calculating the Payments of Winning Bids. (Performed by the ECs Market Using Output of Algorithm 2)

Input: S_{ijkh} : boolean decision variable (from Algorithm 2)
 I : set of available brokers
 J : set of available datacenters
 T : length of lookahead scheduling intervals

Output: p_{ijk} : payment of winning bid Bid_{ijk}

- 1: **for** $i = 0$ to $|I|$ **do**
- 2: **for** $j = 0$ to $|J|$ **do**
- 3: **for** $k = 0$ to $|J|$ **do**
- 4: **for** $h = 0$ to $|T|$ **do**
- 5: **if** ($S_{ijkh} = 1$) **then**
- 6: compute p_{ijk} (Eq. 16)
- 7: **end if**
- 8: **end for**
- 9: **end for**
- 10: **end for**
- 11: **end for**
- 12: **return** p_{ijk}

bounds [46], [47]. Similarly, approximate methods would not provide valid solutions to the VCG auction since they do not guarantee optimality. In order to get a valid solution to the WDP, we implemented it as a Mixed Integer Linear Program (MILP), as shown in Eqs. (6-15), and used the CPLEX solver to solve it to optimality. This WDP implementation can be shown to have an NP-hard time complexity by a reduction from the famous multidimensional knapsack problem [48] to a simplified version of our WDP that only considers one datacenter, a lookahead scheduling horizon of one interval ($T = 1$), no continuity constraints and a unified workloads duration of a single time interval. Such reduction exploits the analogy between a datacenter's computational resources and a knapsack, where scheduled computational workloads cannot exceed the existing datacenter capacity of each resource. Therefore, we accept that the computational complexity of our model is NP-hard; otherwise, that would invalidate the famous $\mathcal{P} \neq \mathcal{NP}$ conjecture [49].

The WDP must be solved once during the winner-determination/resource-allocation phase that determines the auction winners (as shown in Subsection IV-B) and multiple times during the payment-calculation phase that determines the payments required of each winning bidder (as shown in Subsection IV-C). To derive its computational order, we consider a scenario of n cloud brokers, m datacenters, and T lookahead scheduling intervals, where brokers submit bids to migrate workloads to a single datacenter to increase its energy consumption. We assume that all cloud brokers submit bids to migrate workloads of unit duration from $m - 1$ datacenters to a destination datacenter (to maximize the number of received bids and get an upper bound on the number of possible scheduling scenarios), resulting in a total of $n * (m - 1)$ received bids. Since the received bids must be scheduled

within a time horizon of T intervals, the number of all possible scheduling arrangements is $2^{T*n*(m-1)}$. Therefore, the resulting time complexity is $O(2^{T*n*m})$, which is exponential in the number of problem size. Since commercial solvers, such as CPLEX, usually use variants of the Branch-and-Bound algorithm to solve MILPs, their space requirement depends on the branching strategy used to explore the solution space and the number of active tree nodes they keep in memory at any given time [50], [51]. While a worst-case upper bound would require keeping all tree nodes in memory, most commercial solvers would need to store only a subset of the search tree as they normally discard sub-optimal and infeasible branches as soon as they are discovered.

V. PERFORMANCE EVALUATION

This section presents the simulation setup, used parameters, and performance metrics, and discusses our results. We built our simulation model using the Java programming language and used the IBM ILOG CPLEX optimization solver V12.5.1 (via its Concert interface library for Java) to solve the optimization calculations described in the previous section. We measure the performance of our system at different levels of ECs' availability w.r.t the total ECs demanded by the received bids. We also vary the brokers' workload arrival rates to observe the effect of system load on auction results. We use the following metrics to measure our system performance: 1) The effect of using ECs on the grid's balancing time. This is an important metric since power grids operate under strict constraints in terms of balancing time requirements, where a timely response to imbalance contingencies is essential, 2) The sale revenue generated by auctioning ECs compared to selling at regular price, and 3) The brokers' savings as a result of taking part in the ECs' auction and migrating workloads.

A. SIMULATION SETUP

Our simulation model consists of 4 cloud datacenters and 6 cloud brokers, each with randomly chosen cost parameters at the beginning of each simulation run. The time dimension is divided into discrete intervals of equal duration; the system presented in this paper is executed at the beginning of each interval if there are ECs to be auctioned. We simulated our system using 500 independent runs with randomly generated inputs and plotted the averaged results with 95% confidence intervals shown. The ranges for the used parameters are as listed in Table 3. Since the VCG auction charges each winning bidder its social cost as a payment, it is necessary that supply (the amount of auctioned ECs) must always be kept less than demand (the total amount of ECs requested in the received bids) for the VCG auction to collect payments from the bidders. Otherwise, the auction would charge them nothing if their presence does not affect the outcome for other bidders. Hence, we always limit the amount of auctioned ECs to a fraction of the total ECs requested if supply is greater than demand (we experimented with the fractional values of 0.7, 0.8, and 0.9 for this purpose) in order to extract payments from winning bidders. To illustrate, assume that the available

TABLE 3. Simulation Parameters.

Parameter	Range
Broker Parameters	
$CostB^c$	$\in U(80, 120)$ \$/GFLOPS
$CostB^m$	$\in U(80, 120)$ \$/GB
$CostB^s$	$\in U(80, 120)$ \$/GB
Datacenter Parameters	
$CostD^c$	$\in U(20, 40)$ \$/GFLOPS
$CostD^m$	$\in U(20, 40)$ \$/GB
$CostD^s$	$\in U(20, 40)$ \$/GB
$CostD^e$	$\in U(60, 100)$ \$/kWh
$ConvD^c$	$\in U(0.2, 0.5)$
$ConvD^m$	$\in U(0.1, 0.2)$
$ConvD^s$	$\in U(0.1, 0.2)$
$Capacity^c$	500 GFLOPS
$Capacity^m$	500 GB
$Capacity^s$	500 GB
EC	500
Workload Parameters	
w^c	$\in U(2, 5)$ GFLOPS
w^m	$\in U(2, 5)$ GB
w^s	$\in U(2, 5)$ GB
w^d	$\in U(1, 3)$ Intervals

amount of ECs is 1000, while the total amount of ECs requested by the received bids during an iteration is only 100. To avoid giving the ECs to the bidders for free, we limit the amount of auctioned ECs during this iteration to 70, 80, and 90, respectively, according to the different fraction values shown above. The remaining ECs are then auctioned again in the following iteration. Each simulation run starts with 500 ECs available at a randomly selected datacenter. Thus, multiple auctioning iterations (executed at the beginning of each time interval) are normally needed to consume the available ECs. In order to observe the system's performance at different system load levels, we used different brokers' workload arrival rates (10, 20, 30, and 40 requests/scheduling-interval), as shown in the figures.

B. EFFECT OF USING EC AUCTION ON THE GRID BALANCING TIME

Our first performance metric, the effect of using ECs on the grid balancing time (the number of time intervals needed to consume a certain amount of excess energy), shows that selling the excess energy in an auction as ECs can reduce the number of time intervals needed to consume it by as much as 75%. Therefore, shortening the time needed to balance the grid. To illustrate this effect, we plot in Figure 5 the number of time intervals needed to consume a certain amount of excess energy under normal workflow conditions (without using an ECs' auction) and compare it to the case of using an ECs' auction at different levels of ECs' availability and workload arrival rates. The results show that the achievable speed-up in balancing time depends on the system load condition (workload arrival rate) and the amount of ECs made available in the auction. Larger speed-up factors are noticed at lighter system load conditions (compared to the case of not using an ECs' auction) since the ECs' auction helps concentrate brokers' workloads at the destination datacenter. As the system load increases (at higher request arrival rates), the achievable speed-up in balancing time becomes limited by

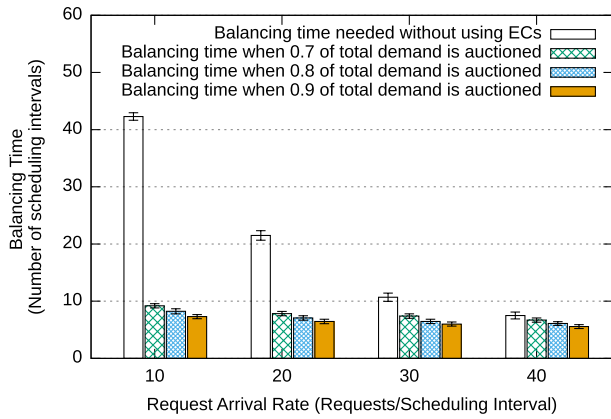


FIGURE 5. Effect of varying the amount of auctioned ECs on balancing time.

the physical capacity of the destination datacenter since not all received bids can be accepted due to capacity constraints (Eqs. 7, 8, and 9). On the other hand, increasing the fraction of auctioned ECs had the effect of shortening the time needed to consume the excess energy since more workload migrations could be accepted.

C. REVENUE GENERATED BY THE EC AUCTION

The second performance metric, the revenue generated by the ECs’ auction, shows how much revenue the seller (the issuing utility) was able to generate by selling the excess energy in an auction, compared to selling at its regular price. Figure 6 shows the regular price of the auctioned excess energy (the bar labeled “Original value”) next to the revenue generated by the ECs’ auction at different levels of ECs’ availability (0.7, 0.8, and 0.9 as discussed above) and at different workload arrival rates. This metric captures the percentage of revenue generated by the ECs’ auction compared to the regular energy price. The results show that selling excess energy in an auction can fetch between 55 to 65% of its regular price. Additionally, reducing the amount of ECs offered in each auction iteration increases the auction revenue due to increased competition for the available ECs.

Cloud brokers always choose the lowest cost datacenter for hosting their workloads. Therefore, they would only migrate their workloads to another (higher-cost) datacenter if lower energy prices at the auction allow them to reduce their cost by migrating. The proposed auction guarantees this to be the case for accepted bids (as shown in Figure 3). This is true since the bid valuations v_{ijk} (shown in Figure 3 and computed using Eq. 5) are always less than the regular energy price at the destination datacenter. Additionally, the actual auction payments p_{ijk} (Eq. 16) are guaranteed not to exceed v_{ijk} . Therefore, the ECs’ auction revenue will always be less than the regular energy price at the destination datacenter, as illustrated by the results in Figure 6. Although power grids would collect only a fraction of the regular price on their excess energy sold in an auction, they would still choose to auction their excess energy to avoid more costly balancing alternatives.

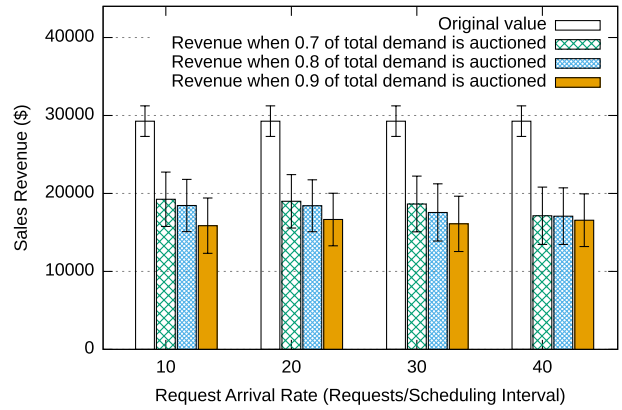


FIGURE 6. Effect of varying the amount of auctioned ECs on total revenue.

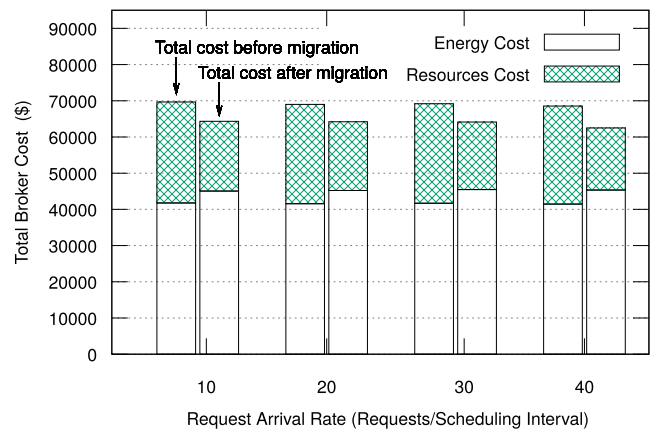


FIGURE 7. Effect of ECs on broker savings.

D. EFFECT OF EC AUCTION ON BROKERS COST

The last performance metric, the effect of ECs on brokers’ cost, captures the reduction in brokers’ cost due to taking part in the auction and migrating workloads. Figure 7 shows the brokers’ cost before and after migration for different workload arrival rates. The results show that auction participants always gain positive utility (i.e., they do not lose money by participating in the auction). The auction mechanism guarantees this since the bid valuations v_{ijk} (shown in Figure 3) guarantee that the total cost after migrating to a destination datacenter never exceeds the original cost at the source datacenter. Additionally, the payment rule (Subsection IV-C) guarantees that winning bidders never pay more than their valuations. Therefore, brokers are always guaranteed not to lose money by joining the auction. While the brokers’ savings are moderate in the shown simulation results (a maximum savings of 10% for the brokers in the considered scenarios), it is worth noting that a bidding broker may always choose to set a minimum threshold of savings for migrating its workload if it so wishes by adjusting its bid value (v_{ijk}) to include this margin.

VI. LIMITATIONS OF THE PROPOSED APPROACH

The proposed system assumes collaboration between the different system entities. Specifically, it assumes that cloud brokers must have access to the energy consumption information of their workloads and the hardware-specific energy consumption parameters of the datacenters, which is not always readily available information from current public cloud datacenters on the market. Due to the unavailability of such datacenters' parameters, we used synthetically generated data (as shown in Table 3) to conduct the system's performance evaluation. While using synthetic data might shift the performance values from those of real production systems, we believe that the general trend of the shown results would still hold since such results are achieved by using the auction mechanism that guarantees positive utility to the participants.

Since the proposed system migrates cloud workloads from multiple source datacenters to a single destination datacenter, we assume that the energy imbalances created at the source datacenters (as a result of workload migrations) are small enough individually and can each be balanced using existing balancing mechanisms of the local balancing areas of each source datacenter. Therefore, based on this assumption, we rule out the possibility of cascading workload migrations.

Since the existing datacenter-based grid balancing approaches discussed in Section-II only consider the case of minimizing datacenters' energy consumption, they are not directly comparable to our approach. Therefore, we show the performance of our approach here compared to the alternative case of not using the ECs' auction. While other non-datacenter-based approaches do exist for balancing excess energy, those are outside the scope of our work since they are not in direct competition with our approach. They can work in parallel with our approach to enhance the grid's balancing capabilities and enable the integration of more RESs' generation units.

VII. CONCLUSION

Grid balancing is essential for the reliable operation of modern utility grids. Datacenters can play an important role in achieving effective grid balancing by providing demand-side flexibility via cloud workload migrations. This flexibility is greatly needed to enable the integration of more RESs into the generation mix of modern grids. While the existing datacenter-based grid-balancing work has focused on minimizing the datacenters' energy consumption (to provide downward flexibility) and has assumed owner-operated datacenters, we used public cloud datacenters in this work as managed loads to provide upward flexibility, which is necessary for balancing excess RESs' energy. We presented a new combined auctioning-scheduling optimization model for increasing the energy consumption at a specific datacenter by a certain amount enough to effectively balance the excess RESs' energy.

We showed via simulation that the proposed solution was effective in ramping up the energy consumption of a target datacenter by allocating its idle resources to minimize the

time needed to consume excess energy by as much as 75%. Plus, we showed that selling excess energy using the proposed auction mechanism did salvage 55%-65% of the original energy cost and provided 10% of cost savings to buyers, thus, provided positive utility to all system participants.

As future work, we are planning several enhancements to the system model, such as including the cost of migration, maximizing the sale revenue of the auction, and targeting the NP-hard complexity of the system. Finally, we are also planning to investigate the use of a decentralized design, such as using a distributed ECs' market and decentralized machine learning approaches.

REFERENCES

- [1] H. L. Ferreira, R. Garde, G. Fulli, W. Kling, and J. P. Lopes, "Characterisation of electrical energy storage technologies," *Energy*, vol. 53, pp. 288–298, May 2013.
- [2] N. Rezaei, A. Ahmadi, S. N. Afifi, A. F. Zobaa, and S. H. E. A. Aleem, "Overview of energy storage technologies," in *Energy Storage at Different Voltage Levels: Technology, Integration, and Market Aspects*, A. F. Zobaa, P. F. Ribeiro, S. H. E. A. Aleem, and S. N. Afifi, Eds. Institution of Engineering and Technology, 2018, pp. 1–29.
- [3] O. M. Babatunde, J. L. Munda, and Y. Hamam, "Power system flexibility: A review," *Energy Rep.*, vol. 6, pp. 101–106, Feb. 2020.
- [4] M. Z. Degefa, I. B. Sperstad, and H. Sæle, "Comprehensive classifications and characterizations of power system flexibility resources," *Electr. Power Syst. Res.*, vol. 194, May 2021, Art. no. 107022.
- [5] L. Hernandez, C. Baladron, J. Aguiar, B. Carro, A. Sanchez-Esguevillas, J. Lloret, and J. Massana, "A survey on electric power demand forecasting: Future trends in smart grids, microgrids and smart buildings," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 3, pp. 1460–1495, 3rd Quart., 2014.
- [6] I. K. Nii, M. Teimeh, O. Nyarko-Boateng, and A. F. Adekoya, "Electricity load forecasting: A systematic review," *J. Electr. Syst. Inf. Technol.*, vol. 7, no. 1, pp. 1–19, Dec. 2020.
- [7] M. F. Anjos and A. J. Conejo, "Unit commitment in electric energy systems," *Found. Trends Electr. Energy Syst.*, vol. 1, no. 4, pp. 220–310, 2017.
- [8] F. Al Farsi, M. Albadi, N. Hosseinzadeh, and A. Al Badi, "Economic dispatch in power systems," in *Proc. IEEE 8th GCC Conf. Exhib.*, Feb. 2015, pp. 1–6.
- [9] K. Ullah, A. Basit, Z. Ullah, S. Aslam, and H. Herodotou, "Automatic generation control strategies in conventional and modern power systems: A comprehensive overview," *Energies*, vol. 14, no. 9, p. 2376, Apr. 2021.
- [10] R. Deng, Z. Yang, M. Chow, and J. Chen, "A survey on demand response in smart grids: Mathematical models and approaches," *IEEE Trans. Ind. Informat.*, vol. 11, no. 3, pp. 570–582, Jun. 2015.
- [11] P. Siano, "Demand response and smart grids—A survey," *Renew. Sustain. Energy Rev.*, vol. 30, no. 2, pp. 461–478, 2014.
- [12] P. Denholm, M. O'Connell, G. Brinkman, and J. Jorgenson, "Overgeneration from solar energy in California: A field guide to the duck chart," Nat. Renew. Energy Lab., Golden, CO, USA, Tech. Rep. NREL/TP-6A20-65023, Nov 2015. [Online]. Available: <https://www.osti.gov/biblio/1226167-overgeneration-from-solar-energy-california-field-guide-duck-chart>
- [13] L. Olmos, M. Rivier, and I. Perez-Arriaga, "Renewable energy and transmission networks," *Green Energy Technol.*, vol. 164, pp. 373–393, Jan. 2015.
- [14] M. Madrigal and S. Stoft, *Transmission Expansion for Renewable Energy Scale-Up: Emerging Lessons and Recommendations*, no. 9375. Washington, DC, USA: The World Bank, Jul. 2012.
- [15] V. Mladenov, V. Chobanov, and A. Georgiev, "Impact of renewable energy sources on power system flexibility requirements," *Energies*, vol. 14, no. 10, p. 2813, May 2021.
- [16] H. Kondziella and T. Bruckner, "Flexibility requirements of renewable energy based electricity systems—A review of research results and methodologies," *Renew. Sustain. Energy Rev.*, vol. 53, pp. 10–22, Jan. 2016.
- [17] N. E. Koltzakis, A. S. Dagoumas, and I. P. Panapakidis, "Impact of the penetration of renewables on flexibility needs," *Energy Policy*, vol. 109, pp. 360–369, Oct. 2017.

- [18] X. Zhang, G. Hug, Z. Kolter, and I. Harjunkoski, "Industrial demand response by steel plants with spinning reserve provision," in *Proc. North Amer. Power Symp. (NAPS)*, Oct. 2015, pp. 1–6.
- [19] V. Lakshmanan, H. Saele, and M. Z. Degefa, "Electric water heater flexibility potential and activation impact in system operator perspective—Norwegian scenario case study," *Energy*, vol. 236, Dec. 2021, Art. no. 121490.
- [20] C. Jennings, "As data center growth surges, so does the energy requirement," *Mil. Eng.*, vol. 109, no. 710, pp. 51–53, 2017.
- [21] H. Rong, H. Zhang, S. Xiao, C. Li, and C. Hu, "Optimizing energy consumption for data centers," *Renew. Sustain. Energy Rev.*, vol. 58, pp. 674–691, May 2016.
- [22] L. A. Barroso and U. Hölzle, "The case for energy-proportional computing," *Computer*, vol. 40, no. 12, pp. 33–37, Dec. 2007.
- [23] M. Lin, A. Wierman, L. L. H. Andrew, and E. Thereska, "Dynamic rightsizing for power-proportional data centers," in *Proc. IEEE INFOCOM*, Apr. 2011, pp. 1098–1106.
- [24] A. Krioukov, P. Mohan, S. Alspaugh, L. Keys, D. Culler, and R. H. Katz, "NapSAC: Design and implementation of a power-proportional web cluster," in *Proc. 1st ACM SIGCOMM Workshop Green Netw.*, Jan. 2010, pp. 15–22.
- [25] Gartner. (2017). *2017 Planning Guide for Cloud Computing*. Accessed: Feb. 2020. [Online]. Available: <https://www.gartner.com/en/documents/3471551>
- [26] G. B. Andresen, R. A. Rodriguez, S. Becker, and M. Greiner, "The potential for arbitrage of wind and solar surplus power in Denmark," *Energy*, vol. 76, pp. 49–58, Nov. 2014.
- [27] Y. V. Makarov, C. Loutan, J. Ma, and P. de Mello, "Operational impacts of wind generation on California power systems," *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 1039–1050, May 2009.
- [28] K. De Vos, "Negative wholesale electricity prices in the German, French and Belgian day-ahead, intra-day and real-time markets," *Electr. J.*, vol. 28, no. 4, pp. 36–50, May 2015.
- [29] R. S. Shannon, "Bringing the customer to the market: A new utility business model," *Electr. J.*, vol. 29, no. 5, pp. 15–18, Jun. 2016.
- [30] M. E. Wainstein and A. G. Bumpus, "Business models as drivers of the low carbon power system transition: A multi-level perspective," *J. Cleaner Prod.*, vol. 126, pp. 572–585, Jul. 2016.
- [31] Y. G. Rebours, D. S. Kirschen, M. Trotignon, and S. Rossignol, "A survey of frequency and voltage control ancillary services—Part II: Economic features," *IEEE Trans. Power Syst.*, vol. 22, no. 1, pp. 358–366, Feb. 2007.
- [32] M. Ghamkhari and H. Mohsenian-Rad, "Data centers to offer ancillary services," in *Proc. IEEE 3rd Int. Conf. Smart Grid Commun. (SmartGridComm)*, Nov. 2012, pp. 436–441.
- [33] H. Chen, A. Coskun, and M. Caramanis, "Real-time power control of data centers for providing regulation service," in *Proc. 52nd IEEE Conf. Decis. Control*, Dec. 2013, pp. 4314–4321.
- [34] H. Nguyen, J. Song, and Z. Han, "Demand side management to reduce peak-to-average ratio using game theory in smart grid," in *Proc. IEEE INFOCOM Workshops*, Mar. 2012, pp. 91–96.
- [35] Z. Liu, I. Liu, S. Low, and A. Wierman, "Pricing data center demand response," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 42, no. 1, pp. 111–123, Jun. 2014.
- [36] X. Cao, J. Zhang, and H. V. Poor, "Data center demand response with on-site renewable generation: A bargaining approach," *IEEE/ACM Trans. Netw.*, vol. 26, no. 6, pp. 2707–2720, Dec. 2018.
- [37] H. Ma, V. Robu, N. Li, and D. Parkes, "Incentivizing reliability in demand-side response," in *Proc. 25th Int. Joint Conf. Artif. Intell. (IJCAI)*, 2016, pp. 352–358.
- [38] H. Wang, J. Huang, X. Lin, and H. Mohsenian-Rad, "Exploring smart grid and data center interactions for electric power load balancing," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 41, no. 3, pp. 89–94, Jan. 2014.
- [39] C. Wang, B. Urganakar, Q. Wang, G. Kesidis, and A. Sivasubramaniam, "Data center power cost optimization via workload modulation," in *Proc. IEEE/ACM 6th Int. Conf. Utility Cloud Comput.*, Dec. 2013, pp. 260–263.
- [40] A. Beloglazov and R. Buyya, "Adaptive threshold-based approach for energy-efficient consolidation of virtual machines in cloud data centers," in *Proc. 8th Int. Workshop Middleware Grids, Clouds e-Sci. (MGC)*. New York, NY, USA: Association for Computing Machinery, 2010, pp. 1–6.
- [41] C. Clark, K. Fraser, S. Hand, J. Hansen, E. Jul, C. Limpach, I. Pratt, and A. Warfield, "Live migration of virtual machines," in *Proc. NSDI*, 2005, pp. 273–286.
- [42] H. R. Varian and C. Harris, "The VCG auction in theory and practice," *Amer. Econ. Rev.*, vol. 104, no. 5, pp. 442–445, May 2014.
- [43] W. Vickrey, "Counterspeculation, auctions, and competitive sealed tenders," *J. Finance*, vol. 16, no. 1, pp. 8–37, 1961.
- [44] S. Dobzinski and N. Nisan, "Limitations of VCG-based mechanisms," *Combinatorica*, vol. 31, no. 4, pp. 379–396, Nov. 2011.
- [45] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [46] E. Chlamtac and M. Tulsiani, "Convex relaxations and integrality gaps," in *Handbook on Semidefinite, Conic and Polynomial Optimization* (International Series in Operations Research & Management Science), M. F. Anjos and J. B. Lasserre, Eds. Boston, MA, USA: Springer, Dec. 2012, pp. 139–169.
- [47] L. Li, *Selected Applications of Convex Optimization* (Springer Optimization and Its Applications). Berlin, Germany: Springer, 2015.
- [48] J. Puchinger, G. R. Raidl, and U. Pferschy, "The multidimensional knapsack problem: Structure and algorithms," *INFORMS J. Comput.*, vol. 22, no. 2, pp. 250–265, May 2010.
- [49] B. Korte and J. Vygen, *Combinatorial Optimization: Theory and Algorithms*. Berlin, Germany: Springer, 2018. [Online]. Available: <https://link.springer.com/book/10.1007/978-3-662-56039-6>
- [50] D. R. Morrison, S. H. Jacobson, J. J. Sauppe, and E. C. Sewell, "Branch-and-bound algorithms: A survey of recent advances in searching, branching, and pruning," *Discrete Optim.*, vol. 19, pp. 79–102, Feb. 2016.
- [51] P. J. Kolesar, "A branch and bound algorithm for the knapsack problem," *Manage. Sci.*, vol. 13, no. 9, pp. 723–735, 1967.



AHMED ABADA received the M.A.Sc. degree in electrical engineering from Carleton University, Ottawa, where he is currently pursuing the Ph.D. degree. His research interests include renewable energy, power grid systems, cloud computing, computer networks, and optimization techniques.



MARC ST-HILAIRE (Senior Member, IEEE) received the Ph.D. degree in computer engineering from Polytechnique Montréal, in 2006. He is currently a Professor with the School of Information Technology with a cross-appointment with the Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada. He is conducting research on various aspects of wired and wireless communication systems. His research interests include network planning and design, network architecture, mobile computing, and cloud computing. With more than 150 publications, his work has been published in several journals and international conferences. Over the years, he has received several awards, including the Carleton Faculty Graduate Mentoring Award, the Carleton Teaching Achievement Award, and several best paper awards. Finally, he is actively involved in the research community. In addition to serving as a member of technical program committees of various conferences, he is equally involved in the organization of several national and international conferences and workshops.



WEI SHI (Member, IEEE) received the Bachelor of Computer Engineering degree from the Harbin Institute of Technology (HIT) and the master's and Ph.D. degrees in computer science from Carleton University, Ottawa, Canada. She is currently a Professor with the School of Information Technology cross appointed to the Department of Systems and Computer Engineering, Faculty of Engineering and Design, Carleton University. She is specialized in the design and analysis of fault tolerance algorithms addressing security issues in distributed environments, such as data-center networks, clouds, mobile agents and actuator systems, wireless sensor networks, as well as critical infrastructures, such as power grids and smart cities. She has also been conducting research in data privacy and big data analytics. She is also a Professional Engineer licensed in Canada.

• • •