Electric Power Systems Research 189 (2020) 106723

Contents lists available at ScienceDirect

Electric Power Systems Research

journal homepage: www.elsevier.com/locate/epsr

Flexibility from networks of data centers: A market clearing formulation with virtual links

Weiqi Zhang[a], Line A. Roald[a], Andrew A. Chien[b,c], John R. Birge[b], Victor M. Zavala[*,a]

[a] University of Wisconsin-Madison, Madison, WI, USA
[b] University of Chicago, Chicago, IL, USA
[c] Argonne National Laboratory, Lemont, IL, USA

ARTICLE INFO

Keywords:
Data centers
Space-time flexibility
Markets

ABSTRACT

Data centers owned and operated by large companies have a high power consumption that is expected to increase in the future. However, the ability to shift computing loads geographically and in time can provide flexibility to the power grid. We introduce the concept of virtual links to capture space-time load flexibility provided by geographically-distributed data centers in market clearing procedures. We show that the virtual link abstraction fits well into existing market clearing frameworks and can help analyze and establish market design properties. This is demonstrated using illustrative case studies.

1. Introduction

Information and computing technologies are the fastest growing uses of electric power (accounting for 8% in 2016 and projected at 13% by 2027), implying that the *dynamics and spatial layouts* of computing loads will increasingly affect power grid operations. At the same time, computing infrastructures are undergoing major structural changes. First, enterprise and government computing functions are being *consolidated into fewer and larger scale facilities*, to support efficient and reliable processing of time-sensitive workloads like search services as well as large-scale workloads that run more effectively on large shared computing facilities [1]. This centralization benefits from economies of scale, such as increased utilization and energy efficiency, but concentrates power loads in a smaller number of locations. According to the Natural Resources Defense Council [2], U.S. data centers in 2013 consumed 90 TWh (the output of 34 power plants with capacities of 500 MW) which is projected to grow to 140 TWh by 2020 (an increase of 55%). To give some perspective on the magnitude of data center loads, we note that large data centers are in the order of 100 MW, with potentially several nearby locations. It has been recently reported that training a single machine learning (ML) model can consume as much as 500 MWh and ML models are documented as growing in size at as rates as high as 10x per year [3].

Second, emergence of "hyperscalars" (e.g., Amazon, Google, Facebook, Microsoft, Alibaba, Tencent) that provide consumer internet and cloud computing services to billions of users has lead to the emergence of networks of data centers. These networks are managed and controlled *collectively* (typically via network operation centers (NOCs)). The consumption of these collections are already large, for example, Google's collection consumed 26 TWh in 2017, and is growing fast [4]. Across these networks, computing tasks can be modulated at multiple timescales (from milliseconds to hours) and can be shifted geographically over long distances (within and beyond the boundaries of independent system operators (ISOs) managing the power grid). This enables *space-time shifting flexibility* of the associated power loads.

Previous research has looked into various ways to harness space-time flexibility from data centers via demand response programs. Data centers are able to provide temporal flexibility (i.e., as a storage device) [5]. In low-to-medium power density data centers, this can be achieved using pre-cooling strategies [6]. In more efficient data centers, strategies to shift loads include server idling, load migration, shutdown and idling of servers and storage clusters, and cooling relative to load reduction [7,8]. Research has also focused on using load shifting and demand response in data centers to enable incorporation of renewable energy sources. Strategies include power cappings to match demand and renewable energy supply [9], shifting loads to reduce peak demand in the face of uncertainty [10], and geographical redistribution to facilitate renewable adoption [11,12]. The effectiveness of shifting computing loads to minimize cost has been demonstrated in [8,13,14], considering shifts between multiple electricity markets [13] and cooperation between data centers [14].

While the potential of harnessing flexibility of data centers in power

* Corresponding author.
E-mail address: victor.zavala@wisc.edu (V.M. Zavala).

https://doi.org/10.1016/j.epsr.2020.106723
Received 3 October 2019; Received in revised form 17 April 2020; Accepted 2 August 2020
Available online 15 August 2020
0378-7796/

grid operations is promising, gathering such flexibility can be challenging due to the complex nature of data center workloads [15]. Therefore, it is necessary to analyze how electricity market designs can influence and incentivize provision of flexibility. The use of online auctioning models to study incentives needed to harness spatial and temporal flexibility from data centers has been studied in [16,17]. A Nash bargaining formulation is employed to analyze interactions between data centers and load serving entities [18] and between data centers and tenants [19].

In this work, we propose market clearing models that capture space-time flexibility provided by load shifting and migration in data centers. Shifting can be achieved by a company owning a set of geographically distributed data centers, which allows management of where and when to satisfy the loads. Specifically, the company can choose to shift the workload to servers at another location or to delay the workload to later times. We note that previous work has studied models and mechanisms for the provision of load flexibility from a data center perspective. Limited studies have analyzed load shifting mechanisms and models from a systems-wide perspective (ISO perspective). For instance, recent work has revealed that shiftable loads can help absorb stranded power and control power flows in the network [20,21]. Here, we study the load shifting problem from a coordinated market clearing perspective and present formulations that accommodate space-time shifting flexibility. We take the perspective of an ISO, which seeks to clear the market by using demand and supply bidding information in a coordinated manner. Recent examples [22–26] seek to analyze how changes in market design (i.e., in the bidding process and clearing formulation) can influence dispatch and price behavior and can benefit ISO operations and market participants.

The main contributions of this paper are as follows: We introduce the notion of *virtual links*, which are (non-physical) pathways that shift loads in space (by reallocating computing loads to other geographical location) and time (by delaying a computing load). The network of virtual links acts as an additional infrastructure layer on top of the existing transmission grid. We show that this paradigm is intuitive and compatible with existing market clearing procedures (in which flexibility is provided by a physical transmission network and generator dynamics). The proposed framework also reveals which information should be provided by data centers in the bidding process and provides insights into conditions that will incentivize data centers to provide flexibility. We also show that virtual links provide a convenient framework to establish pricing and social welfare properties and to analyze complex space-time behavior. Specifically, by means of case studies, we demonstrate that data center flexibility leads to higher social welfare and to higher total loads delivered to data centers. Moreover, we also demonstrate that virtual links lead to local marginal prices (LMPs) that tend to become more spatially and temporally homogeneous because virtual links help relieve transmission network congestion and generator ramping limits.

The paper first presents the market clearing setting in Section 2. We start from the traditional setting without any virtual links, and then introduce both spatial and temporal shifts. Section 3 then demonstrates the benefits on different case studies, while Section 4 summarizes and concludes.

2. Market clearing setting

The market setting presented here builds on formulations presented in [23,24], but is adapted to capture space-time load shifting flexibility. This first part of the section discusses a baseline setting where no load shifts are captured in the market clearing procedure (shifts are introduced in Section 2.1).

We begin by considering a time-independent network (of the physical electric grid) with a set of geographical nodes $\mathcal{N} = \{n_0, n_1, ..., n_N\}$, suppliers $\mathcal{G}$, loads (data centers) $\mathcal{D}$, and physical transmission lines $\mathcal{L}$. We also consider a set of entities or stakeholders that own (or operate) loads $\mathcal{E}$. Each generator $i \in \mathcal{G}$ is connected to node $n(i) \in \mathcal{N}$, has a total flow $p_i \in \mathbb{R}_+$, maximum supply capacity $\bar{p}_i \in \mathbb{R}_+$, and bidding cost $\alpha_i^p \in \mathbb{R}_+$. For convenience, we define the set of generators connected to node n as $\mathcal{G}_n = \{i \in \mathcal{G} | n(i) = n\}$.

Each load $j \in \mathcal{D}$ is connected to node $n(j) \in \mathcal{N}$, has a served load $d_j \in \mathbb{R}_+$, requested load $\bar{d}_j \in \mathbb{R}_+$, and bidding value $\alpha_j^d \in \mathbb{R}_+$. The served loads refer to the outcome of the market clearing. We define the set of loads connected to node n as $\mathcal{D}_n \subseteq \mathcal{D}$. Each load has an associated owning/operating entity $e(j) \in \mathcal{E}$ and we define the set of loads owned by entity e as $\mathcal{D}_e \subseteq \mathcal{D}$. Note that the notion of set of loads can be generalized to include both data center and non-datacenter loads. Each link $l \in \mathcal{L}$ has an associated receiving (end) node rec(l), sending (source) node snd(l), physical (bidirectional) power flow $f_l \in \mathbb{R}$, maximum capacity $\bar{f}_l$, and bidding cost α_l^f. We use $\mathcal{L}_n^{in}$ to denote the set of lines with flow that enter node n and $\mathcal{L}_n^{out}$ to denote the set of lines with flow leaving node n. We define the market clearing prices (locational marginal prices-LMPs) at node n as $\pi_n \in \mathbb{R}$.

Based on the setting described, a basic market clearing formulation (solved by the ISO) takes the form:

$$\max_{(d,p,f)\in C} \phi := \sum_{j\in\mathcal{D}} \alpha_j^d d_j - \sum_{i\in\mathcal{G}} \alpha_i^p p_i - \sum_{l\in\mathcal{L}} \alpha_l^f |f_l| \tag{1a}$$

$$\text{s.t.} \sum_{l\in\mathcal{L}_n^{in}} f_l + \sum_{i\in\mathcal{G}_n} p_i = \sum_{l\in\mathcal{L}_n^{out}} f_l + \sum_{j\in\mathcal{D}_n} d_j,\ n \in \mathcal{N} \tag{1b}$$

$$f_l = B_l(\theta_{\text{snd}(l)} - \theta_{\text{rec}(l)}), \quad l \in \mathcal{L} \tag{1c}$$

where $C = \{(d, p, f) | 0 \le p_i \le \bar{p}_i,\ 0 \le d_j \le \bar{d}_j, |f_l| \le \bar{f}_l\}$ denotes capacity constraints of generators, loads, and flows. The objective function (1a) is the social welfare, which seeks to maximize value of the total load served while minimizing operating costs associated with generation and physical transmission. Constraints (1b) enforce power conservation at each node. Eq. (1c) denote DC flow constraints, where θ_n is the voltage angle at node n and B_l is the susceptance of line l. Note that the proposed model is different from standard DC-OPF formulations in that it explicitly considers transmission cost terms in the social welfare (which can be interpreted as operating costs of transmission lines). The standard DC-OPF can be recovered by setting the transmission bid cost to zero. This observation will be relevant when exploring properties of virtual load shifting.

A solution of the clearing problem can also be found by solving the Lagrangian dual problem,

$$\max_{\pi} \min_{(d,p,f)\in\mathcal{F}} \mathcal{L}(d, p, f) \tag{2}$$

where the partial Lagrange function is given by

$$\begin{aligned}\mathcal{L}(d, p, f) &= \sum_{i\in\mathcal{G}} \alpha_i^p p_i + \sum_{l\in\mathcal{L}} \alpha_l^f |f_l| - \sum_{j\in\mathcal{D}} \alpha_j^d d_j \\ &- \sum_{n\in\mathcal{N}} \pi_n \left(\sum_{l\in\mathcal{L}_n^{in}} f_l + \sum_{i\in\mathcal{G}_n} p_i - \sum_{l\in\mathcal{L}_n^{out}} f_l - \sum_{j\in\mathcal{D}_n} d_j \right) \\ &= \sum_{j\in\mathcal{D}} \phi_j + \sum_{i\in\mathcal{G}} \phi_i + \sum_{\ell\in\mathcal{L}} \phi_\ell.\end{aligned} \tag{3}$$

Here, the feasible set $\mathcal{F}$ includes the set C and the set of feasible injections induced by the DC constraints [24]. This indicates that the clearing problem finds dual variables (LMPs) π_n for the balance constraints that maximize the profit functions for generators $\phi_i := (\pi_{n(i)} - \alpha_i^s)p_i$, loads $\phi_j := (\alpha_j^d - \pi_{n(j)})d_j$, and transmission lines $\phi_l := (|\Delta\pi_l| - \alpha_l^f)|f_l|$ (where $\Delta\pi_l := \pi_{\text{rec}(l)} - \pi_{\text{snd}(l)}$ are the nodal price differences) [27]. Intuitively, these profit functions evaluate how much welfare each stakeholder gains (or loses if negative) from the market clearing outcome (in addition to their individual bids). If the set $\mathcal{F}$ accepts zero as a feasible clearing solution (all clearing quantities are zero), one can establish that the profits of all players are non-zero [23,24,27].

2.1. Market clearing with spatial shifts

We now introduce the concept of *virtual links* to capture *spatial load shifting* flexibility provided by data centers. Consider that an entity e operating a set of data centers $\mathcal{D}_e$ at nodes $\mathcal{N}_e$ offers flexibility to the ISO by allowing to shift part of a load d_j, $j \in \mathcal{D}_e$ from the base node $n(j)$ (which we refer to as the *hub node*) to a set of alternate nodes $\mathcal{N}_e$. As mentioned in the introduction, this can be done by using computing load migration, where part of the computing loads at one cluster is migrated to another cluster. The shifted computing load is received at the source cluster, but executed at the destination cluster. This hence shifts the demand of electricity from one cluster to the other. Since the load shifting requires at least two data centers at different locations, this type of flexibility can be offered by entities that own and operate multiple data centers at different geographical locations (e.g., a computational task and associated power load can be executed at different locations). We assume that each data center consumes the same amount of energy for a given computing task. Migrating computing tasks might, however, be associated with a reduction in the quality of service to customers of the data center. Also, this may require the datacenter operator to keep multiple copies of customer data, or could increase the time required to serve a certain customer request. Therefore, the datacenter operators might require a payment for shifting loads in space (they provide a service).

To capture load shifting flexibility in the market clearing, we use *virtual links* as non-physical pathways to quantify this kind of load shifting capability between pairs of clusters located at different nodes, in units of electric power consumed by the shifted jobs. In this way, we can think of data centers owned by on entity as an extra layer of transmission network that does not have to adhere to Kirchoff's laws. We will show later that this alternative form of transmission (provided by data center flexibility) helps clear more loads and reduce electricity price volatility by overcoming physical constraints.

We define a set of virtual links for the load requesting entity $e \in \mathcal{E}_d$ as $\mathcal{V}_e$. Each link $v \in \mathcal{V}_e$ has capacity to send a portion of the served load d_j (denoted as $\delta_v \in \mathbb{R}$) from node $n(j)$ to node $n(j')$ with $n(j)$, $n(j') \in \mathcal{N}_e$ (i.e., $\mathrm{snd}(v) = n(j)$ and $\mathrm{rec}(v) = n(j')$). To avoid degeneracy, we assume that load cannot be sent and received at the same node (no virtual link v exists with $\mathrm{snd}(v) = n(j)$ and $\mathrm{rec}(v) = n(j)$). A virtual link provides a *non-physical* pathway for the shifted load δ_v (in the sense that the link does not carry power). As a computing task is shifted to a data center at another node, its associated load is *physically* cleared at the destination node. Thus, the total load that is (physically) absorbed at node n is given by:

$$\hat{d}_n = \sum_{j \in \mathcal{D}_n} d_i + \sum_{v \in \mathcal{V}_n^{in}} \delta_v - \sum_{v \in \mathcal{V}_n^{out}} \delta_v$$

and the total load absorbed at node $n(j)$ needs to satisfy the constraint $0 \le \hat{d}_{n(j)} \le d_n^{\max}$, where $d_n^{\max}$ is the capacity of the data center at node n. Here, $\mathcal{V}_n^{in}$, $\mathcal{V}_n^{out}$ are the sets of virtual links that enter and leave node n, respectively. If $\delta_v = 0$ for all virtual flows v entering and leaving the node $n(j)$, the original load d_j is served at the hub node $n(j)$ and thus $\hat{d}_{n(j)} = \sum_{j \in \mathcal{D}_n} d_j$. On the other hand, if a portion of the load is shifted then this will be physically absorbed at another node. The power conservation equation is thus modified as:

$$\begin{aligned} \sum_{l \in \mathcal{L}_n^{in}} f_l + \sum_{i \in \mathcal{G}_n} p_i &= \sum_{l \in \mathcal{L}_n^{out}} f_l + \hat{d}_n \\ &= \sum_{l \in \mathcal{L}_n^{out}} f_l + \sum_{j \in \mathcal{D}_n} d_j + \sum_{v \in \mathcal{V}_n^{in}} \delta_v - \sum_{v \in \mathcal{V}_n^{out}} \delta_v. \end{aligned}$$

Virtual links allow the ISO to shift loads between nodes to overcome physical constraints associated with generation and transmission. Moreover, we will see that virtual links provide alternative pathways that can be exploited to increase economic efficiency and to mitigate price volatility. We also observe that adding virtual links is *mathematically equivalent* to the introduction of transmission lines with the exception that the flows do not have to adhere to Kirchoff's laws. Note also that, by definition, if an entity e does not own/operate multiple loads (i.e., the sets $\mathcal{D}_e$ and $\mathcal{N}_e$ are singletons) then it cannot offer spatial shifting flexibility. Furthermore, entities that choose to shift power across the network may incur a cost due to increased latency. This provides motivation for thinking of data centers owned by one entity as a network with edges represented by virtual links. Based on these observations, it does make sense to assume that virtual links are offered as products (services) in the market clearing framework. Accordingly, we consider the formulation:

$$\begin{aligned} \max_{(d,p,f,\delta) \in C} \quad & \textstyle\sum_{j \in \mathcal{D}} \alpha_j^d d_j - \sum_{i \in \mathcal{G}} \alpha_i^p p_i - \sum_{l \in \mathcal{L}} \alpha_l^f |f_l| - \sum_{v \in \mathcal{V}} \alpha_v^\delta |\delta_v| \\ \text{s.t.} \quad & \textstyle\sum_{l \in \mathcal{L}_n^{in}} f_l + \sum_{i \in \mathcal{G}_n} p_i + \sum_{v \in \mathcal{V}_n^{out}} \delta_v \\ & \quad = \textstyle\sum_{l \in \mathcal{L}_n^{out}} f_l + \sum_{j \in \mathcal{D}} d_j + \sum_{v \in \mathcal{V}_n^{in}} \delta_v, \quad n \in \mathcal{N} \\ & f_l = B_l(\theta_{\mathrm{snd}(l)} - \theta_{\mathrm{rec}(l)}), \quad l \in \mathcal{L} \end{aligned}$$

where $C = \{(d, p, f, \delta) | 0 \le p_i \le \bar{p}_i,\ 0 \le d_j \le \bar{d}_j,\ 0 \le \hat{d}_n \le d_n^{\max},\ |f_l| \le \bar{f}_l\}$. The formulation can also impose maximum allowable shifts between nodes of the form $\underline{\delta}_l \le \delta_l \le \bar{\delta}_l$. These constraints can help capture technical limitations of data centers or asymmetries in allowable shifts (e.g., virtual flows can only move in one direction).

One can directly use duality concepts (similar to Eqs. (2)–(3)) to show that the above formulation maximizes the profit functions for generators, loads, and transmission links (as in a standard clearing formulation) and also for virtual links $\phi_v := (|\Delta\pi_v| - \alpha_v^\delta)|\delta_v|$, where $\Delta\pi_v := \pi_{\mathrm{rec}(v)} - \pi_{\mathrm{snd}(v)}$, and α_v^δ is the cost associated with shifting a load. We thus have that virtual shifting capacity provides an additional revenue stream to market participants. This also offers mechanism to hedge against bidding risk because offering the opportunity to shift loads will more likely result in more load being served (i.e., the ISO can reconfigure loads to maximize total load served), compared to the case in which an entity requires load at different nodes but does not offer the opportunity to reconfigure them. This behavior is illustrated in Fig. 1.

2.2. Market clearing with temporal shifts

Although the concept of virtual links naturally arises from the way spatial load shifting is achieved, we can generalize this concept to represent temporal flexibility. Through methods like idling servers, data centers are able to pause the computing jobs when a demand response event starts and restart them afterwards [7]. Therefore, we can use a virtual link that branches out from one time point to another in the future to represent the workload delay *within* a data center. Similar to the case of spatial flexibility, the virtual link quantifies how much workload is delayed in units of energy consumed, but with a different assumption that no load is dropped due to time delays. We use a cost term in the objective function to capture decrease in quality of service caused by workload delay.

We now show how virtual flows can be used capture *temporal load shifting* flexibility. For simplicity, we consider a single spatial location and define the lexicographic set $\mathcal{T} := \{t_1, \ldots, t_T\}$ to represent a sequence of

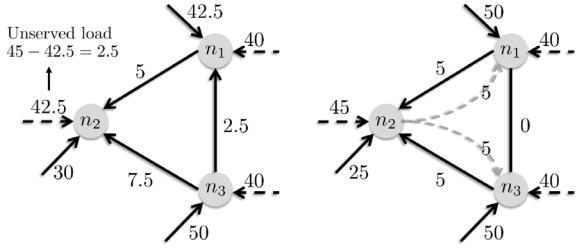


Fig. 1. The clearing outcome of 3-bus case study scenarios 1 (left) and 4 (right). The black dashed lines denote loads, black solid lines supplies or power transmission (if between two nodes). Grey dashed curves denote virtual links.

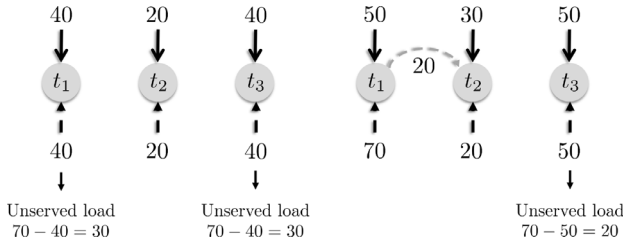


Fig. 2. The clearing outcome of 5-time case study scenarios 1 (left) and 3 (right). Only the first 3 time points are shown. The black dashed lines denote loads, black solid lines supplies or power transmission (if between two nodes). Grey dashed curves denote virtual links.

locations (nodes) in time. The interpretation of time points as nodes will become relevant when capturing space-time behavior. Now consider that the entity e requests loads $d_j,\ j \in \mathcal{D}_e$ to be delivered at times $t_j \in \mathcal{T}$ and thus these nodes are interpreted as the hub nodes. If the loads cannot be met at the hub nodes, the data center offers flexibility to shift a fraction of these loads to other available times, which is captured in the node set $\mathcal{T}_e \subseteq \mathcal{T}$. As before, this is done by using virtual links $\mathcal{V}_e$ that connect the hub nodes to the set of available nodes $\mathcal{T}_e$. This behavior is illustrated in Fig. 2.

We observe that virtual flows can also be used to capture storage dynamics. To illustrate this, we assume that the firm requests loads d_t at nodes $t \in \mathcal{T}$ and define the set of virtual links $\mathcal{V}_e$: $=\{v_1, ..., v_{T-1}\}$ with element v_l constructed such that $\text{snd}(v_l) = t_l$ and $\text{rec}(v_l) = t_{l+1}$ for $l = 0,\ ..., T-1$. We note that load balances at the time nodes can be written as:

$$\delta_t = \delta_{t-1} + u_t,\ \ t = t_1,\ ..., t_T$$

with $\delta_{t_1} = u_{t_1}$, $\delta_T = 0$ and u_t: $=d_t - \hat{d}_t$. In other words, the virtual flows δ_t act as storage that carries over unmet demand u_t over future times. Imposing constraints on virtual flows allow us to control the load carryover and the bidding cost for the virtual flows can be used to capture the fact that a shift might become increasingly expensive as one moves forward in time (e.g., delaying a computation load becomes increasingly expensive as the delay increases). The storage interpretation assumes that carryover only occurs forward in time but we note that virtual flows can be used for arbitrary shifts in time to allow for strategic reallocation during the clearing procedure.

2.3. Market clearing with spatio-temporal shifts

The virtual flows in time and space can be combined to enable a straightforward generalization to capture spatio-temporal shifting flexibility that facilitates analysis and interpretation of clearing formulations. We define a set of spatial nodes $\mathcal{N}$ and a set of temporal nodes as $\mathcal{T}$. In a standard clearing formulation, the supply, load, and transmission flows at time $t \in \mathcal{T}$ are defined as $p_{i,t}$, $d_{j,t}$, and $f_{l,t}$. In a market setting that captures space-time flexibility, an entity e requests loads $d_{j,t}$ at hub nodes $(n(j),\ t) \in \mathcal{N} \times \mathcal{T}$ but also offers the possibility to shift the loads to a set of space-time nodes $\mathcal{N}_e \times \mathcal{T}$. We define the set of virtual links $\mathcal{V}_e$, where each link v sends a virtual flow $\delta_v \in \mathbb{R}$ (fraction of the load $d_{j,t}$) from the space-time hub node $(n(j),\ t)$ to another space-time node $(n(j'),\ t')$; in other words, we have that $\text{snd}(v) = (n(j),\ t)$ and $\text{rec}(v) = (n(j'),\ t')$. As before, to avoid degeneracy, we assume that load cannot be sent and received at the same space-time node (no virtual link v exists with $\text{snd}(v) = (n(j),\ t)$ and $\text{rec}(v) = (n(j),\ t)$). The total load served at space-time location $(n,\ t)$ is given by:

$$\hat{d}_{n,t} = \sum_{j \in \mathcal{D}_n} d_{j,t} + \sum_{v \in \mathcal{V}^{in}_{n,t}} \delta_v - \sum_{v \in \mathcal{V}^{out}_{n,t}} \delta_v$$

The market clearing formulation takes the form:

$$\begin{aligned}
\max_{(d,p,f,\delta) \in C} \quad & \sum_{t \in \mathcal{T}} \sum_{j \in \mathcal{D}} \alpha^d_{j,t} d_{j,t} - \sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{G}} \alpha^p_{i,t} p_{i,t} \\
& - \sum_{t \in \mathcal{T}} \sum_{l \in \mathcal{L}} \alpha^f_{l,t} |f_{l,t}| - \sum_{v \in \mathcal{V}} \alpha^\delta_v |\delta_v| \\
\text{s.t.} \quad & \sum_{l \in \mathcal{L}^{in}_n} f_{l,t} + \sum_{i \in \mathcal{G}_n} p_{i,t} + \sum_{v \in \mathcal{V}^{out}_{n,t}} \delta_v \\
& = \sum_{l \in \mathcal{L}^{out}_n} f_{l,t} + \sum_{j \in \mathcal{D}_n} d_{j,t} + \sum_{v \in \mathcal{V}^{in}_{n,t}} \delta_v, \quad (n,\ t) \in \mathcal{N} \times \mathcal{T} \\
& f_{l,t} = B_l(\theta_{\text{snd}(l),t} - \theta_{\text{rec}(l),t}), \qquad (l,\ t) \in \mathcal{L} \times \mathcal{T}
\end{aligned}$$

where

$C = \{(d, p, f, \delta) \mid 0 \le p_{i,t} \le \bar{p}_{i,t},\ 0 \le d_{j,t} \le \bar{d}_{j,t},\ 0 \le \hat{d}_{n,t} \le d^{\max}_{n,t},\ |f_{l,t}$. The $| \le \bar{f}_{l,t},\ |p_{i,t+1} - p_{i,t}| \le \Delta\bar{p}_i\}$

latter constraint in C captures the ramping constraints for generators, which may create time congestion (analogous to transmission congestion in the spatial domain). We will see that temporal virtual shifts allow us to alleviate congestion generated from ramping constraints while spatial shifts alleviate congestion generated from transmission constraints.

3. Case studies

In this section we demonstrate various benefits of using virtual links. We present small case studies to illustrate the key results and a large study to show that the results are scalable. The codes and data for all case studies presented in this section can be found in https://github.com/zavalab/JuliaBox/tree/master/PSCC2020.

3.1. Spatial shifts in a 3-bus network

We consider a network with three spatial nodes and one time node (i.e., we assume there is no temporal virtual shift). Each node contains one supply and one data center and each pair of nodes contains one physical transmission line and one spatial virtual link. We assume that a single entity operates the three data centers. The generation capacities are set to $\bar{p} = \{50,\ 30,\ 50\}$, load capacities to $\bar{d} = \{40,\ 45,\ 40\}$, transmission capacity to $\bar{f} = \{5,\ 10,\ 10\}$, generation bidding costs to $\alpha^p = \{10,\ 20,\ 10\}$, load bidding prices to $\alpha^d = \{40,\ 30,\ 40\}$, and transmission costs for the links to $\alpha^f = \{2,\ 2,\ 2\}$. For each transmission line l, we set $B_l = 0.5$. The system is designed in a way that nodes n_1 and n_3 have excess in supply while n_2 has an excess in load. The system has three virtual links $\mathcal{V} = \{(1,\ 2),\ (1,\ 3),\ (2,\ 3)\}$ available to shift load. We solve the market clearing problem under six different scenarios that account for increasing levels of spatial flexibility.

The scenarios and results are summarized in Table 1. Here, $\bar{\delta}$ and α^δ_v denote the maximum shifting capacity and bidding cost for all virtual links. Scenario 1 corresponds to the case in which no spatial flexibility is available from data centers. Scenarios 2 to 5 gradually increase shifting capacities. In scenarios 6 and 7 we use the same shift capacities as scenario 4 but we decrease the shifting cost to show interplay with transmission costs.

The results show that the social welfare increases as data centers offer more capacity to shift loads, due to an increase in the amount of load served and the more efficient use of generation. Interestingly, in scenarios 3 and 4, transmission flows decrease as more load is served (even though the bid cost of physical transmission is lower than that of virtual links). This is because physical transmission line flows are limited by the line constraints and the Kirchoff laws, but can be overcome using virtual links. Scenario 6 shows that, when the shift cost is lower than the transmission cost, no physical transmission is used at all and instead loads are reconfigured using virtual shifts to obtain the highest social welfare. This illustrates how virtual shifting can provide an economic alternative to transmission and that this can create incentives for the provision of load flexibility by data centers.

Table 1 also shows that the nodal prices become more homogeneous

Table 1
Results for three-bus network with temporal shifting flexibility.

Scenario	$\bar{\delta}$ (MWh)	α_v^{δ} ($/MWh)	ϕ ($)	π ($/MWh)	d (MWh)	p (MWh)	f (MWh)	δ (MWh)
1	[0,0,0]	3	2920	[10,30,18]	[40,42.5,40]	[42.5,30,50]	[5, − 2.5, −7.5]	[0,0,0]
2	[5,0,0]	3	2980	[10,20,13]	[40,45,40]	[47.5,27.5,45]	[5, − 2.5, −7.5]	[−5,0,0]
3	[15,0,0]	3	2997.5	[17,20,16.5]	[40,45,40]	[50,25,50]	[5, − 2.5, −7.5]	[−7.5,0,0]
4	[15,0,15]	3	3000	[17,20,17]	[40,45,40]	[50,25,50]	[5,0, − 5]	[−5,0,5]
5	[100,100,100]	3	3000	[17,20,17]	[40,45,40]	[50,25,50]	[5,0, − 5]	[−5,0,5]
6	[15,0,15]	1	3030	[19,20,19]	[40,45,40]	[50,25,50]	[0,0,0]	[−10,0,10]
7	[15,0,15]	0	3050	[20,20,20]	[40,45,40]	[50,25,50]	[0,0,0]	[−10,0,10]

Table 2
Results for one-bus network with temporal shifting flexibility.

Scenario	$\bar{\delta}$ (MWh)	ϕ ($)	π ($/MWh)	d (MWh)	p (MWh)	δ (MWh)
1	[0,0,0,0]	5200	[30, − 30,40,15,20]	[40,20,40,40,40]	[40,20,40,40,40]	[0,0,0,0]
2	[10,0,0,0]	5770	[30,0,40,15,20]	[60,20,50,40,40]	[50,30,50,40,40]	[10,0,0,0]
3	[21,0,0,0]	5840	[23,20,40,15,20]	[70,20,50,40,40]	[50,40,50,40,40]	[20,0,0,0]
4	[21,20,0,0]	5840	[23,20,40,15,20]	[70,20,50,40,40]	[50,40,50,40,40]	[20,0,0,0]
5	[21,0,21,0]	6060	[23,20,40,37,20]	[70,20,60,40,40]	[50,40,50,50,40]	[20,0,10,0]
6	[21,0,21,10]	6200	[23,20,26,23,20]	[70,20,70,40,40]	[50,40,50,50,50]	[20,0,20,10]
7	[100,100,100,100]	6200	[23,20,26,23,20]	[70,20,70,40,40]	[50,40,50,50,50]	[20,0,20,10]

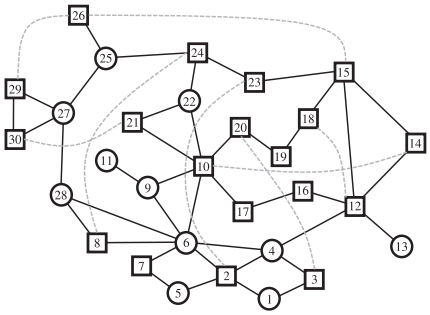
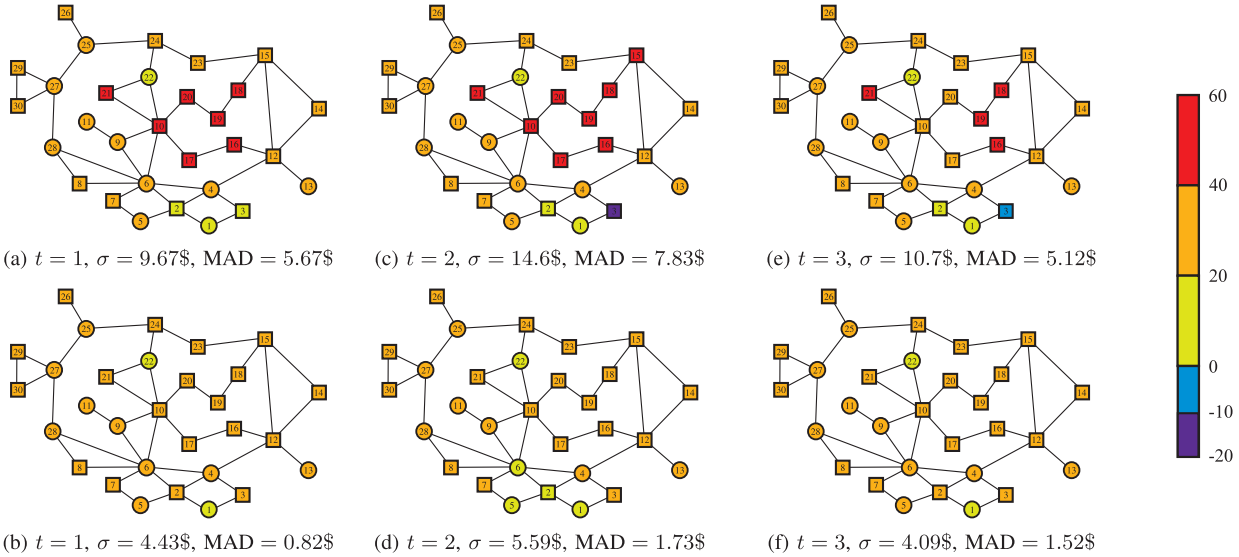


Fig. 3. Schematic of IEEE 30-bus system showing transmission links (solid) and spatial virtual links (dotted). Nodes are either attached to loads (square) or not (circle). Only a fraction of spatial virtual links are shown for clarity.

across the network as the system gains more spatial shifting flexibility. For scenarios 1 to 4, the range of the prices shrinks as the data centers provide more shifting capacity. We also observe that the nodal price distribution exhibits convergence as flexibility increases. Specifically, scenarios 4 and 5 show that, once the prices converge, additional shifting flexibility does not fully homogenize prices. This is related to another observation that, at the limiting solution, the price difference between nodes is the cost of spatial shift. This result can be established from duality and is analogous to the well-known result that nodal price differences are bounded by transmission costs [27]. The virtual shift cost can be understood as the minimum incentive the market should provide to compensate for the cost of the spatial shift. We also observe that in scenario 7, with zero shifting costs, the nodal prices can be made fully homogeneous as shifting capacity is increased.

3.2. Temporal flexibility in a one-bus network

We now consider a one-node network with one supplier and one data center that offers temporal flexibility over a time horizon of five points. The virtual links are defined to create a storage-like system: at each time node the data center receives loads that are shifted from the previous time and any unserved load is carried over. As boundary



(a) $t = 1$, $\sigma = 9.67\$$, MAD $= 5.67\$$ (c) $t = 2$, $\sigma = 14.6\$$, MAD $= 7.83\$$ (e) $t = 3$, $\sigma = 10.7\$$, MAD $= 5.12\$$

(b) $t = 1$, $\sigma = 4.43\$$, MAD $= 0.82\$$ (d) $t = 2$, $\sigma = 5.59\$$, MAD $= 1.73\$$ (f) $t = 3$, $\sigma = 4.09\$$, MAD $= 1.52\$$

Fig. 4. Space-time prices for IEEE 30-bus system without (top row) and with (bottom row) virtual links. The LMP value of each node is denoted by its color. The variance (σ) and mean absolute deviation (MAD) values of LMP at each time point are shown in the subtitles.

Table 3
Generator $\bar{p}$ Data (MWh).

Node #	$t = 1$	$t = 2$	$t = 3$
1	83.3955	68.5275	83.9986
2	82.9331	82.5598	79.3669
13	51.3812	37.6343	35.7812
22	49.7154	57.6143	43.359
23	32.4078	27.6153	19.6892
27	54.0729	66.569	55.738

Table 4
Transmission line data.

Line l	B_l	$\bar{f}_l$ (MWh)
(1,2)	15.0	5.6921
(1,3)	4.9223	17.6822
(2,4)	5.2308	16.225
(2,5)	4.7059	18.554
(2,6)	5.0	17.0763
(3,4)	23.5294	3.7108
(4,6)	23.5294	3.7108
(4,12)	3.8462	23.4
(5,7)	7.1006	11.7
(6,7)	10.9589	7.6896
(6,8)	23.5294	3.7108
(6,9)	4.7619	18.9
(6,10)	1.7857	17.8571
(6,28)	15.0	5.6921
(8,28)	4.5872	18.7926
(9,10)	9.0909	9.9
(9,11)	4.7619	18.9
(10,17)	10.9589	7.6896
(10,20)	4.023	20.5626
(10,21)	12.069	6.8542
(10,22)	5.4745	14.8977
(12,13)	7.1429	12.6
(12,14)	3.1707	25.7721
(12,15)	5.9633	13.2883
(12,16)	4.158	19.7385
(14,15)	2.2624	22.6244
(15,18)	3.6364	22.1371
(15,23)	4.0	20.1246
(16,17)	4.4706	18.554
(18,19)	6.3415	12.886
(19,20)	12.069	6.8542
(21,22)	40.0	2.0125
(22,24)	3.8462	19.47
(23,24)	3.0067	26.97
(24,25)	2.2759	22.7586
(25,26)	1.8366	18.3664
(25,27)	3.7367	21.3359
(27,28)	2.5	25.0
(27,29)	1.8683	18.6833
(27,30)	1.2976	12.9758
(29,30)	1.7301	17.301

conditions we have that, at $t = t_1$, the data center receives no shifted loads while, at $t = t_5$, the data center cannot shift any load to the next time. Within the time window, the load capacity and bid costs of loads and supplies change with time. The system thus have $T = 5$ time nodes and $T - 1$ virtual links $\mathcal{V}$: ={(1, 2), (2, 3), (3, 4), (4, 5)}. The generation capacities are set to $\bar{p} = \{50, 50, 50, 50, 50\}$, load capacities to $\bar{d} = \{70, 20, 70, 40, 40\}$, generation bidding costs to $\alpha^p = \{10, 20, 10, 15, 20\}$, and load bidding prices to $\alpha^d = \{30, 60, 40, 50, 45\}$. We fix the bidding cost for virtual links as $\alpha^\delta = \{3, 3, 3, 3\}$. For ramp limits, we set $\bar{\Delta p} = 20$.

The problem setup follows the sketch in Fig. 2. We use seven scenarios of different temporal shift capacities to demonstrate important properties, presented in Table 2. The results are analogous to those observed in the spatial shifting case (this highlights how virtual links facilitate treating space-time dimensions in a unified framework). Specifically, the social welfare and the total amount of delivered loads increase with increasing shift capacity. The price variance over the time nodes becomes narrower as shifting capacity is offered. In the limit of high shifting capacity, prices converge and the differences between nodes are bounded by the shifting cost. This illustrates how properties induced by spatial and temporal virtual links are analogous. We note that scenario 1 has a negative LMP caused by the ramping limit, which is relieved in later scenarios by virtual links. Another interesting observation (also shown in Fig. 2) is that for scenarios 1 to 3, even if only a virtual link between t_1 and t_2 is added, the amount of load cleared at t_3 increases. This shows how temporal flexibility is able to relieve ramping constraints. A district property of the temporal shifts used here, however, is that their effect on the price gaps is unidirectional. In particular, due to the fact that loads can only be shifted forward in time, temporal shifts of loads can take advantage of a potentially lower price in the

Table 5
Data center $\bar{d}$ data (MWh).

Node #	$t = 1$	$t = 2$	$t = 3$
2	23.3782	20.7436	25.9256
3	3.0080	0.0	5.8536
4	5.1955	7.61508	12.9122
7	21.827	22.8585	27.9395
8	44.4138	41.3557	47.3557
10	4.4034	0.0	9.536
12	5.4430	7.3421	11.0886
14	5.8174	10.4422	7.1592
15	14.284	12.6535	0.3244
16	5.3331	2.28527	4.25272
17	8.9176	8.4928	24.2347
18	4.8253	2.0689	3.0487
19	17.2053	0.0	2.9341
20	5.2518	0.0	0.0
21	24.2327	21.3645	26.164
23	11.5308	6.3614	4.6527
24	12.2126	3.3929	19.6662
26	0.0	12.5725	7.3736
29	1.7507	9.2522	8.2738
30	13.6553	4.6596	8.8905

future for more revenue, but not vice versa. This property is illustrated in scenarios 3 and 4, where additional flexibility in link $v = (2, 3)$ does not change the solution since the price at node 2 is higher than that at node 3.

3.3. Space-time flexibility in IEEE 30-bus network

We now consider the provision of space-time flexibility in a test case based on the IEEE 30-bus power network. While the topology and location of loads and generators remains the same as in the original test case [28], we have adjusted the load, generation, and line susceptances to reflect a system with variable generation capacity (due to, e.g., renewable energy variability) and a large number of data centers. The data for this modified system is described in the appendix. For simplicity, we consider a time window of three units. Each pair of data centers has a spatial virtual link which allows them to exchange load spatially within each time period. Each data center is able to defer its own load until a later point in time, i.e., it has a temporal virtual link from each time point to the next time point. For each virtual link, we impose an upper bound for the spatial and temporal shits of at most 10 units of power. The cost of virtual shifts are reflected by a bid cost of 0.0001 for each spatial link and a bid cost of 3 for each temporal link. A schematic of the network with spatial virtual links is shown in Fig. 3. The optimization problem is solved with and without the data center flexibility.

Fig. 4 shows the space-time price distributions. For the case without virtual links, the heat maps (Fig. 4a to e) show a high level of spatial and temporal price volatility. Specifically, we can observe multiple congestion nodes (nodes with prices greater than or equal to 40) and several nodes with low prices and the price range increases at certain times. This spatial and temporal price volatility is relieved with the addition of virtual shifts. Specifically, as shown in Fig. 4b to f, at each time point the price range shrinks. Moreover, we have found that the social welfare increases by 100% (from $5,209 to $11,217) by incorporating virtual links.

From Fig. 4 we observe clustering behavior for the space-time prices. For example, in the case without virtual links, the prices of nodes 10, 17, 20 remain at the same level, while the price of nodes 2 is noticeably lower throughout the entire time window. When virtual links are added the prices become more homogeneous and follow a similar trend. We note that this homogenization behavior is not limited to cluster of nodes that are close to each other geographically. For instance, nodes 21 and 30 have an associated data center and therefore are connected by a spatial virtual link. Due to the added shifting flexibility, they also exhibit the price homogenization, even though they are far apart. This illustrates how virtual links can help overcome geographical barriers associated with transmission network topology.

In this study we also observed the emergence of negative prices. For the case without data center flexibility, node 3 has negative prices at time 2 and 3. In the cases of negative prices, electricity flows from node 1 to node 3, and from node 3 to node 4. This is caused by transmission constraints. When the system incorporates virtual shifting, space-time price volatility reduces and negative prices disappear.

4. Conclusions and future work

We presented the concept of virtual links as a way to capture space-time load shifting flexibility of data centers in market clearing. We show that virtual links are mathematically equivalent to transmission links, which facilitate the analysis and interpretation of the clearing quantities and prices. Moreover, we show that virtual links can be used to capture space time behavior in a systematic manner. Case studies show that virtual shifts lead to higher social welfare, higher load served, and mitigation of space-time price volatility. Our results also highlight that virtual links provide an additional source of revenue for data centers and thus create incentives to provide flexibility. In future work we will refine our clearing model to capture various constraints that limit the flexibility that data centers can offer in practice. Also, a more refined evaluation of flexibility costs is needed to account for specific penalties (e.g., quality of service reduction).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We acknowledge support from the U.S. NSF under award 1832208. The fourth author acknowledges support from the University of Chicago Booth School of Business.

Appendix A.

Generator, data center and transmission line data for the test case based on the modified IEEE 30-bus case study are provided in Tables 3–5, respectively.

References

[1] L.A. Barroso, U. Hölzle, P. Ranganathan, The datacenter as a computer: designing warehouse-scale machines, Synth. Lect. Comput. Archit. 13 (3) (2018) i–189.
[2] P. Delforge, J. Whitney, Data center efficiency assessment-scaling up energy efficiency across the data center industry: Evaluating key drivers and barriers, (2014). Natural Resources Defense Council
[3] D. Amodei, D. Hernandez, Ai and compute, 2018.
[4] E. Dreyfuss, How google keeps its power-hungry operations carbon neutral, 2018.
[5] Z. Liu, I. Liu, S. Low, A. Wierman, Pricing data center demand response, ACM SIGMETRICS Perform. Eval. Rev. 42 (1) (2014) 111–123, https://doi.org/10.1145/2637364.2592004.
[6] M. Lukawski, J.W. Tester, M.C. Moore, P. Krol, C.L. Anderson, Demand response for reducing coincident peak loads in data centers, Proc. of the 52nd Hawaii Int. Conf. on System Sciences, (2019).
[7] G. Ghatikar, V. Ganti, N. Matson, M.A. Piette, Demand response opportunities and enabling technologies for data centers: findings from field studies(2012). 10.2172/1174175.
[8] J. Li, Z. Bao, Z. Li, Modeling demand response capability by internet data centers processing batch computing jobs, IEEE Trans. Smart Grid 6 (2) (2014) 737–747.
[9] D. Gmach, J. Rolia, C. Bash, Y. Chen, T. Christian, A. Shah, R. Sharma, Z. Wang, Capacity planning and power management to exploit sustainable energy, 2010 International Conference on Network and Service Management, IEEE, 2010, pp. 96–103.
[10] Z. Liu, A. Wierman, Y. Chen, B. Razon, N. Chen, Data center demand response: avoiding the coincident peak via workload shifting and local generation, Perform. Eval. 70 (10) (2013) 770–791.
[11] H. Wang, Z. Ye, Renewable energy-aware demand response for distributed data centers in smart grid, 2016 IEEE Green Energy and Systems Conference (IGSEC), IEEE, 2016, pp. 1–8.
[12] Z. Liu, M. Lin, A. Wierman, S.H. Low, L.L. Andrew, Geographical load balancing with renewables, ACM SIGMETRICS Perform. Eval. Rev. 39 (3) (2011) 62–66.
[13] L. Rao, X. Liu, L. Xie, W. Liu, Minimizing electricity cost: optimization of distributed internet data centers in a multi-electricity-market environment, 2010 Proc. IEEE INFOCOM, (2010), pp. 1–9.
[14] L. Rao, X. Liu, M.D. Ilic, J. Liu, Distributed coordination of internet data centers under multiregional electricity markets, Proc. IEEE 100 (1) (2011) 269–282.
[15] A. Wierman, Z. Liu, I. Liu, H. Mohsenian-Rad, Opportunities and challenges for data center demand response, International Green Computing Conference, IEEE, 2014, pp. 1–10.
[16] R. Zhou, Z. Li, C. Wu, An online emergency demand response mechanism for cloud computing, ACM Trans. Model. Perform. Eval. Comput. Syst. 3 (1) (2018) 5:1–5:25.
[17] Q. Sun, S. Ren, C. Wu, Z. Li, an online incentive mechanism for emergency demand response in geo-distributed colocation data centers, Proc. of the 7th Int. Conf. on Future Energy Systems, ACM, 2016.
[18] X. Cao, J. Zhang, H.V. Poor, Data center demand response with on-site renewable generation: a bargaining approach, IEEE/ACM Trans. Netw. 26 (6) (2018) 2707–2720.
[19] Y. Guo, H. Li, M. Pan, Colocation data center demand response using nash bargaining theory, IEEE Trans. Smart Grid 9 (5) (2017) 4017–4026.
[20] K. Kim, F. Yang, V.M. Zavala, A.A. Chien, Data centers as dispatchable loads to harness stranded power, IEEE Trans. Sustain. Energy 8 (1) (2016) 208–218.
[21] F. Yang, A.A. Chien, ZCCloud: exploring wasted green power for high-performance computing, 2016 IEEE International Parallel and Distributed Processing Symposium, (2016), pp. 1051–1060.
[22] P.R. Gribik, D. Chatterjee, N. Navid, L. Zhang, Dealing with uncertainty in dispatching and pricing in power markets, 2011 IEEE Power and Energy Society General Meeting, (2011), pp. 1–6.
[23] V.M. Zavala, K. Kim, M. Anitescu, J. Birge, A stochastic electricity market clearing formulation with consistent pricing properties, Oper. Res. 65 (3) (2017) 557–576.
[24] G. Pritchard, G. Zakeri, A. Philpott, A single-settlement, energy-only electric power market for unpredictable and intermittent participants, Oper. Res. 58 (4-NaN-2) (2010) 1210–1219.
[25] M. Carrion, J. Arroyo, A computationally efficient mixed-integer linear formulation for the thermal unit commitment problem, IEEE Trans. Power Syst. 21 (3) (2006) 1371–1378.
[26] F. Bouffard, F.D. Galiana, A.J. Conejo, Market-clearing with stochastic security-part i: formulation, IEEE Trans. Power Syst. 20 (4) (2005) 1818–1826.
[27] A.M. Sampat, Y. Hu, M. Sharara, H. Aguirre-Villegas, G. Ruiz-Mercado, R.A. Larson, V.M. Zavala, Coordinated management of organic waste and derived products, Comput. Chem. Eng. (2019).
[28] R.D. Zimmerman, C.E. Murillo-Sánchez, R.J. Thomas, Matpower: steady-state operations, planning, and analysis tools for power systems research and education, IEEE Trans. Power Syst. 26 (1) (2010) 12–19.