Applied Energy 229 (2018) 69–79

Contents lists available at ScienceDirect

Applied Energy

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

Data centers as a source of dynamic flexibility in smart girds

Ilari Alaperä[a,*], Samuli Honkapuro[b], Janne Paananen[c]

[a] Fortum Power and Heat Oy, Keilalahdentie 2-4, 021 Espoo, Finland
[b] Lappeenranta University of Technology (LUT), Skinnarilankatu 34, 53850 Lappeenranta, Finland
[c] Eaton EMEA, Koskelontie 15, 02920 Espoo, Finland

HIGHLIGHTS

- Data centers have a lot of excess capacity due to their redundant design.
- This capacity can be used to provide ancillary services, such as primary regulation.
- UPS systems are technically and economically viable in primary regulation.
- Additional stress to battery systems is within battery specifications.
- Participation to primary regulation can create significant revenue for data centers.

ARTICLE INFO

Keywords:
Demand response
Primary frequency regulation
Data center
UPS
Battery
Smart grid

ABSTRACT

Data centers have a significant potential to become a major source of flexibility in smart girds. They consume currently roughly 3% of all the electricity produced globally and are expected to only increase their consumption as the world becomes more connected and digitalized.

As data centers are required to operate without any interruptions, they use power protection systems and energy storages. This paper investigates the technical and economic feasibility of dual-purposing these power protection systems, the uninterruptible power supplies, and their batteries in data centers to perform primary frequency regulation services. While the topic of data centers and demand response has been extensively covered in the current scientific literature, the focus has been on the demand response enabled by server workload shifting or hardware-enabled peak shaving. Based on an extensive literature review, there is a knowledge gap in the literature concerning primary frequency regulation and dynamic response enabled by modern power electronics systems in data centers. In this paper, this knowledge gap is bridged by suggesting a novel approach of taking advantage of the bidirectional operations capabilities of the uninterruptible power supply systems, thereby enabling them to provide dynamic power response from their battery systems.

The feasibility of this approach is examined with the proposed method, which includes (1) an analysis of the required energy for primary regulation and the availability of this energy in a typical data center, (2) a simulation of activation events and their impact on the service life of the battery systems, (3) reaction speed and reliability considerations of the operations, and (4) an economic feasibility and balancing market analysis.

The results show that as primary frequency regulation is an energy nonintensive service and data center battery systems are by design oversized for redundancy reasons, typical data centers have more than ample amounts of energy to participate in the primary regulation without jeopardizing their own processes. The results also show that by maintaining reasonable levels of usage, the battery systems can be operated within their specifications, and the demand response operations will not cause premature aging of the battery systems. The reaction speed of the power electronics is found to be very high and easily meet the current market requirements.

While the achievable revenue from the primary regulation service is small compared for example with the electricity costs of the data centers, it is still significant as there is little to no impact on the daily business of the data centers.

* Corresponding author.
E-mail address: ilari.alapera@fortum.com (I. Alaperä).

https://doi.org/10.1016/j.apenergy.2018.07.056
Received 27 April 2018; Received in revised form 28 June 2018; Accepted 13 July 2018
0306-2619/

1. Introduction

As the amount of renewable and intermittent energy increases in electrical power systems globally, also the need for flexibility increases [1]. This flexibility has been traditionally provided by large power plants, but now, as an increasing proportion of production is becoming difficult or even impossible to adjust, new sources of flexibility are needed. Attention is turning to flexible consumption, and the topic of demand response has been addressed in multiple scientific publications. For example, Muhssin et al. and Xu et al. studied the usage of aggregated household consumption assets, such as refrigerators and heat pumps, as sources of ancillary services in [2,3], and Jia et al. investigated the usage of electric vehicles (EVs) in frequency regulation in [4]. Batteries and energy storages and their feasibility in grid support applications, such as primary regulation, have been extensively studied both in academic literature and commercial demonstrations. For example, Zheng et al. analyzed the economic and environmental benefits of different dispatch strategies of a large number of residential energy storage systems in [5], Shi et al. analyzed multipurpose usage of battery systems for peak shaving and frequency regulation in [6], and Cheng et al. and Brivio et al. covered the combined use cases of primary frequency regulation and energy arbitrage in [7,8]. The demand response potential of power-intensive industries has also been studied lately; Otashu et al. proposed a metric of analyzing the available load reduction in the industry in [9], and Ramin et al. presented a case example of a metal casting process in [10].

Data centers are the power-intensive industry of the modern age. They are highly redundant digital factories and among the largest energy consumers in the world, and their power consumption is expected to increase significantly over the following decades. This development is driven by increasing digitalization and a growing amount of data being transferred and processed [11,12]. To achieve the high uptime requirements, data centers are designed with significant amounts of inbuilt flexibility in the form of electrochemical storage (i.e. batteries) and redundant power electronic systems. These design choices and significant power consumption make data centers attractive candidates for demand response participation.

Currently, there are few data centers participating in grid support activities, and while some companies have already begun to commercialize these activities [13,14], the majority of the data center flexibility potential still remains an untapped resource. The participation of data centers in demand response has also been covered in the scientific literature. For example, Mamum et al. addressed the topic in [15,16], where demand response was studied from the perspective of performing peak-shaving with Li-ion batteries, while Li et al. focused on demand response enabled by IT load shifting in [17]. However, as the extensive literature review (presented below) shows, there is a knowledge gap in the current literature concerning data center participation in primary frequency regulation. In particular, the key questions still unanswered are: how data center power protection systems could enable dynamic regulation, and what is the technical and economic feasibility of such an approach.

This paper presents a novel way for the data centers to participate in grid support (specifically in primary frequency regulation) by actively using their uninterruptible power supply (UPS) systems and batteries to balance the grid, instead of shedding their loads by off-gridding their systems, or performing workload shifting. The technical and economic feasibility of this approach is analyzed methodologically from several viewpoints.

The main research questions of this paper are: could a data center with a battery system perform primary frequency response in an economically feasible way without significant risks to their primary business, and how much additional stress would be exerted on the existing batteries during these grid support operations.

The structure of the paper is the following: The results of a literature review are presented in the second section. The third section introduces primary frequency regulation. The fourth section explains typical UPS topologies and their inherent excess capacities deployed in data centers through selected example configurations. The fifth section explains the key differences between the commonly applied method of providing grid services and the proposed approach. The sixth section investigates the technical feasibility of performing dynamic upwards regulation with an UPS system. The seventh section presents results from a frequency analysis and a simulation model to estimate the additional stress exerted on the UPS systems and their batteries while participating in primary regulation, and references it to typical battery cycle life characteristics. The eighth section provides discussion, and conclusions are drawn in the final section.

2. Literature review

The subject of demand response (DR) in the data center space is extensively covered in the current scientific literature. The research can be divided into two main categories; (1) DR enabled by “IT knobs” (basically workload management of the servers) and (2) DR by data center hardware (e.g., UPS systems, air conditioning, generators, and additional on-site generation). By far, the majority of the research focuses on the first case, that is, server management, and specifically, how to enable implicit demand response. The term ‘implicit demand response’ refers to (1) optimizing the electricity consumption of a data center for example by peak shaving to reduce grid connection costs, (2) limiting server power consumption during periods of high electricity prices, or (3) spatial workload shifting to gain savings from regional price differences. Li et al. [17] modeled the effects of spatial and temporal spreading of IT workloads to take advantage of electricity price differences between different price regions and times of the day. Similar research was presented by Ruddy et al. in [18], where they introduce a methodology for shifting global demand and calculate the resulting cost and CO_2 emissions savings and potential DR revenue from capacity payments in the Irish electricity markets. Further, in [19], Liu et al. proposed several algorithms for optimizing electricity cost and the usage of renewable energy for data center operations.

A more extensively studied subtopic has been the role and potential of multitenant or colocation (colo) data centers, especially in emergency demand response (a type of ancillary service, where independent system operators (ISOs) contract resources to respond to their DR signals, dispatched when a grid power balance is in jeopardy). Tran et al. have investigated the subject and related topics in multiple publications; in [20], they presented a simulation model and related results for cost optimizing EDR activations in a mixed-use building with data center (server) loads (workload management), HVAC systems, and backup generators. In their other publications they have studied how to incentivize the colo tenants to participate in the EDR by proposing different schemes and analyzing their convergence rates [21,22]. Similar research was presented by Kishwar et al. in [23] and Sun et al. in [24]. Guo et al. also examined how to incentivize colo tenants’ participation in the EDR and applied the Nash bargaining theory to coordinate the tenants’ participation and revenue gain in [25]. Again, Zhan et al. [26] proposed (and mathematically proved) a pricing model for colo data center operators that would include a reward component for tenants with flexible processes. The colo operator would benefit from reduced grid tariffs (resulting from peak load reduction). There are also several conference publications on the topic of using server management for DR, each with a slightly different focus. Wang et al. presented a DR framework model (server workload management) in [27], and further, they introduced an electricity cost optimization algorithm in [28], which is much similar to the work presented by Baharm et al. in [29,30].

In addition to the above-mentioned work related to server load management, there are several extensive papers that address the topic of using hardware systems in data centers to perform demand response operations. The most relevant research with respect to the scope of this