Virtualizing Balancing Power: An Energy-Aware Load Dispatcher for Cloud Computing

by

Gilbert Fridgen, Robert Keller, Markus Thimmel[1], Lars Wederhake

presented at: Energy Informatics & Management 2015 (EIM), Rotterdam, Netherlands, June 2015

[1] At the time of writing this paper, Markus Thimmel was a Student Assistant at the Research Center Finance & Information Management.

Virtualizing Balancing Power: An Energy-Aware Load Dispatcher for Cloud Computing

Abstract

Balancing mechanisms assure grid stability. Especially well-suited for balancing purposes are large-scale storage facilities (SFs). However, the potential for these is in major parts set by geographic realities. On a transnational level, offering that potential to regions in need of balancing power (BP) does not often appear to be economically viable - an issue that is frequently related to the construction of power lines. Thus, in this article, we illustrate an early version of a design artifact giving remote balancing mechanisms access to a local BP market without deploying power lines: utilization of data centers (DCs) is typically very low (30-40%) representing a cheap source of demand flexibility. We thus let one DC participate in an existing BP market while tying a second to a remote balancing mechanism. By doing so, the design artifact enables both load and BP to flow seamlessly between distinct power markets contributing to grid stability and efficient utilization of balancing mechanisms. Within this extended summary, we perform a preliminary evaluation of the design artifact based on real-world data.

Motivation

Cloud Computing is the basis for a growing number of information services (Armbrust et al., 2010) such as collaboration tools, or big data applications. The cloud computing paradigm emphasizes economies of scale through centrality (Heilig & Voß, 2014). Therefore, it is based on only a few but massive, geographically distributed DCs (Zhang, Wang, & Wang, 2011). Although cloud computing is arguably a more energy-efficient approach to providing information services (Baliga, Ayre, Hinton, & Tucker, 2011), DCs, yet, consume enormous quantities of electricity. All US-based DCs combined contribute some 2% to the country's total electricity consumption (Koomey, 2011). Looking ahead, the electricity demand for cloud infrastructures is projected to soar up by at least 60% by 2020 (Cook, 2014).

Being very specific to electricity markets, demand and supply are tightly coupled by that they have to match in every instant. Considering the ever-growing energy generation from wind and photovoltaic (PV), the shrinking potential for balancing through dynamically ramping up supply side resources is challenging grid stability. During solar eclipse in Europe on March 20th, 2015, this threat has become very real to Germany as worlds' most reliant country on PV (Stetz et al., 2015). Thus, additionally examining the potential of demand side resources for balancing the grid has become an active research field (Falvo, Graditi, & Siano, 2014; Goddard, Klose, & Backhaus, 2014; Kirschen, 2003). By far the vast majority of today's approaches are variants of time-based load shifting, i.e. load is put forward or delayed in time.

Another approach to balancing local power grids is to spatially shift load. Typically, this is done by im-/exporting load through power lines. The exchange, in general, is favorable for both importers and exporters of power for two reasons: first, excess power in one place can neutralize a deficit in another, and second, some markets can provision flexibility more cost-efficiently than others (Van Hulle et al., 2010). Three components determine a markets' efficiency in provisioning flexibility: its mix of energy generation, the potential for adjusting demand, and efficient SFs like (pumped) hydropower plants. The potential for the latter is heavily influenced by geographical realities. Therefore, it is imperative to inter-connect markets in order to reap the benefits from alleviating market inefficiencies. However, many power line construction projects fail for reasons of excessive initial costs (Kishore & Singal, 2014), insecure return on investment (Buijs, Bekaert, Cole, Van Hertem, & Belmans, 2011), storms of protests by local citizens (Lütticke, 2014), and transmission losses.

Meanwhile, serving the power system as a reserve capacity for both supply (positive) and load (negative), BP is continuing to be essential for achieving grid stability in many electricity markets within Europe and abroad (Vandezande, Meeus, Belmans, Saguan, & Glachant, 2010). A BP market coordinates the cost-efficient provision of both positive and negative BP within its market region.

However, if it was possible to actually transfer an energy-consuming set of activities instantly to a different market, BP services can be transferred to regions with a lower potential for balancing mechanisms avoiding issues relating to power lines. Presumably, only information goods/services perfectly fulfill that requirement. They come at very low, still exponentially decreasing transaction costs (Mack, 2011; Moore, 1965). Consequently, a distributed DC service provision could be a mechanism to tender flexibility in markets where it is comparatively more expensive. While this benefits the linked electricity markets, it also opens an additional stream of revenue for DC operators by tendering BP.

Several approaches already investigate the potential of geographically distributed DCs participating in balancing programs by shifting workload. Ghatikar, Ganti, Matson, and Piette (2012) demonstrate by a field study, that it is technical feasible to shift workload, including already commenced jobs, from one DC to another within a few minutes. Wang, Huang, Lin, and Mohsenian-Rad (2014) analyze the interactions between the power grid operator and the operator of geographically distributed DCs. Therefore, the authors study the impact of workload shifting on the power grid. Berl, Klingert, Beck, and de Meer (2013) and Basmadjian, Niedermeier, Lovasz, De Meer, and Klingert (2013) introduce Green Supply Demand Agreements between an electricity provider and a DC operator. These can be part of an electricity tariff for participating in balancing programs, e.g. by migrating workload. However, such novel tariffs are not available yet. Rao, Liu, Xie, and Liu (2012) develop an algorithm to minimize the electricity costs for distributed DCs according to electricity price differences on various spot markets. Furthermore, Chiu, Stewart, and McManus (2012) propose a concept for grid balancing by shifting workload between geographically distributed DCs pursuant to local real-time price signals. Though, the required price signals need for further research.

However, none of these articles studies the concept of transferring the need for BP to distant markets through workload dispatching between geographically distributed DCs considering a real-world BP market. The underlying market design applies to e.g. California, Australia, and Germany (NERC, 2011). Therefore, we believe to carry on the scientific discourse by answering the following research question:

Is it possible to increase the potential for grid balancing among distant power grids by operating two geographically distributed DCs, participating in a real-world BP market, and thereby increase the profitability for the DC operator?

In order to answer this research question, we build in the following a workload dispatching algorithm as an design artifact (Gregor & Hevner, 2013; Hevner, March, Park, & Ram, 2004).

Setting

We consider a setting with two DCs managed by a single operator. These DCs, which can deliver the same cloud services within the same processing-time, are separated by distinct BP market areas. Furthermore, one DC participates in its local BP market, the other collaborates with a SF. The capacity of that SF is assumed to be unlimited. Modern DCs are energy proportional, meaning that its electricity consumption roughly linearly grows in utilization (Beloglazov, Abawajy, & Buyya, 2012). Hence, a DC operator is able to in-/decrease a DC's electricity consumption by shifting utilization from or to another DC. This shift of utilization is implemented by dispatching incoming workload to the DCs and is solely motivated by the retrieved BP. We assume that the individual jobs of the workload are mutually independent of one another, i.e. they do not aggregate to bulks. Thus, the bandwidth is not used for transferring commenced jobs. Furthermore, idle servers are switched off (Beloglazov et al., 2012).

The provided BP must be readily available during the entire period covered by the bid. Therefore, the DC operator should always just bid as much BP as can be safely delivered. In this vein, it seems appropriate to assume that the workload is sufficiently available to deliver the provided BP. Thereby, we imply that the capacity of both DCs is sufficient to provide the agreed amount BP for provision. This assumption seems to be realistic, since "even in virtualized resource pools, utilization is typically below 40%" (Blagodurov et al., 2013), and there are approaches to forecast workload peaks (e.g. Jheng, Tseng, Chao, & Chou, 2014).

The deferrable workload equivalent to the provided positive BP must initially be held in the DC participating in the BP market (Figure 1: DC1) and the deferrable workload equivalent to the provided negative BP must initially be held in the other DC (Figure 1: DC2). Thus, the quantity of electricity purchased from electricity provider 1 correspond to the provided positive BP and the quantity purchased from electricity provider 2 correspond to the provided negative BP.

In each period, the algorithm performs the following case distinction (Figure 1):

Case 1) If the retrieved positive (negative) BP increases (decreases) compared to the prior period, sufficient workload is dispatched to DC2 to decrease the electricity consumption of DC1

Case 2) If the retrieved positive (negative) BP decreases (increases) compared to the prior period, sufficient workload is dispatched to DC1 to increase the electricity consumption of DC1

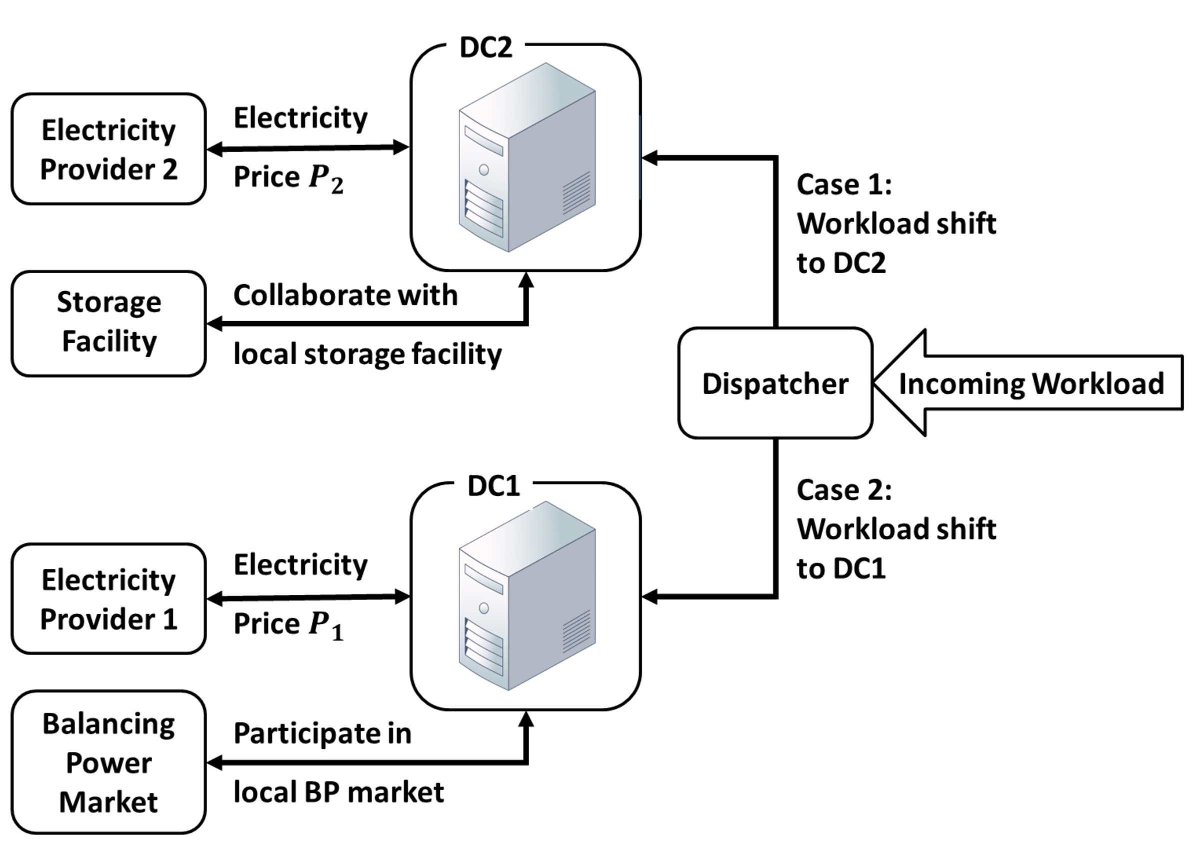

Figure 1 Linking two power grids by dispatching workload

The electricity surplus resulting from decreasing the utilization of DC2 according to the delivered negative BP is used to charge the SF. The deficit of electricity resulting from increasing the utilization of DC2 according to the delivered positive BP is compensated by discharging this SF. Therefore, the electricity demand purchased from the electricity providers in both locations remains constant and the volatility caused by balancing the grid through DC1 is not transmitted to electricity provider 2.

The DC operator utilizes the difference between the compensation for providing balancing services in the market with higher balancing costs and the costs for utilizing balancing services in the other market. However, the shift of the utilization from one DC location to another results in shifting costs. These costs are caused by the fact that the servers in both locations are consuming electricity simultaneously during the booting- shifting- and shutdown-processes.

Evaluation

To evaluate the economic potential of the introduced design artifact, we have to identify the earnings potential for a DC operator providing BP in one location, the shifting costs, and the remuneration for the SF operator in the other location. As a first step evaluation, we investigate the earnings potential for participating in an existing BP market and the shifting costs. The earnings potential minus the shifting costs can be interpreted as an upper bound for the remuneration of the SF operator. This remuneration for balancing the distant grid must exceed the potential revenue of the

SF operator for delivering the same BP services within the local market, as otherwise, he will not collaborate with the DC operator. Moreover, we only consider negative BP. Positive BP and the determination of the potential revenue are subject to further research.

We set up a scenario with one DC in Germany, participating in the German BP market, and another in Norway, as a presumably cheaper source of flexibility. In that scenario, our algorithm is used to deliver the retrieved BP in Germany by dispatching workload. Furthermore, we set up a comparative scenario, within that the DC operator is not participating in the BP market. The base load, consisting of the DC's non-deferrable energy demand, is the same for both scenarios. Therefore, we only include the electricity costs for the bidden BP in our evaluation.

In the scenario with the consideration of our algorithm, the overall electricity costs consist of the remuneration for providing BP, the electricity consumption costs in both locations, and the shifting costs. In the comparative scenario, the DC operator is not participating in a BP market, and thus, the whole deferrable workload is processed by the DC with the lowest electricity price. In our setting, this is DC2 in Norway (IEA, 2014). Consequently, we compare our algorithm to the most favorable comparative scenario.

We implement a heuristic based on real-world data of the German BP market ("regelleistung.net," 2015) and 5 MW of negative BP offered in each tender in the timeframe of one year. We determine an upper bound for the overall electricity costs by an optimal bidding strategy. Therefore, we bid in each tender period the prices resulting in the lowest overall electricity costs in that period. Via a second bidding strategy, we forecast these optimal prices to get a lower bound. We derived a cost-saving potential of approximately 80% (40%) as an upper (lower) bound. Actually, the results might tend towards the upper bound as there will be more elaborated bidding strategies than those in our simple forecast model.

Assuming that the remuneration for the SF operator in Norway is not exceeding the obtained savings from provision BP in Germany, the algorithm results in an economic benefit. This benefit is incentivizing a DC operator for activating SFs in Norway and providing access to the German power market.

Conclusion

In our paper, we introduce a scenario, in that a DC operator participates in a real-world BP market. Furthermore, we implemented the algorithm and two bidding strategies to illustrate the enormous earnings potential.

Regarding our further research, the next step is to evaluate the costs for considering a SF. Moreover, we will investigate the potential for a DC operator to provide positive BP. Finally, we will improve our lower bound bidding strategy in order to even exceed our results.

References

Armbrust, M., Stoica, I., Zaharia, M., Fox, A., Griffith, R., Joseph, A. D., … Rabkin, A. (2010). A view of cloud computing. *Communications of the ACM*, *53*(4), 50. doi:10.1145/1721654.1721672

Baliga, J., Ayre, R. W. A., Hinton, K., & Tucker, R. S. (2011). Green Cloud Computing: Balancing Energy in Processing, Storage, and Transport. *Proceedings of the IEEE*, *99*(1), 149–167. doi:10.1109/JPROC.2010.2060451

Basmadjian, R., Niedermeier, F., Lovasz, G., De Meer, H., & Klingert, S. (2013). GreenSDAs leveraging power adaption collaboration between energy provider and data centres. In *Sustainable Internet and ICT for Sustainability (SustainIT)* (pp. 1–9). IEEE. doi:10.1109/SustainIT.2013.6685195

Beloglazov, A., Abawajy, J., & Buyya, R. (2012). Energy-aware resource allocation heuristics for efficient management of data centers for Cloud computing. *Future Generation Computer Systems*, *28*(5), 755–768. doi:10.1016/j.future.2011.04.017

Berl, A., Klingert, S., Beck, M. T., & de Meer, H. (2013). Integrating data centres into demand-response management: A local case study. In *IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society* (pp. 4762–4767). IEEE. doi:10.1109/IECON.2013.6699905

Blagodurov, S., Gmach, D., Arlitt, M., Chen, Y., Hyser, C., & Fedorova, A. (2013). Maximizing server utilization while meeting critical SLAs via weight-based collocation management. In *2013 IFIP/IEEE International Symposium on Integrated Network Management* (pp. 277–285). IEEE.

Buijs, P., Bekaert, D., Cole, S., Van Hertem, D., & Belmans, R. (2011). Transmission investment problems in Europe: Going beyond standard solutions. *Energy Policy*, *39*(3), 1794–1801. doi:10.1016/j.enpol.2011.01.012

Chiu, D., Stewart, C., & McManus, B. (2012). Electric grid balancing through lowcost workload migration. *ACM SIGMETRICS Performance Evaluation Review*, *40*(3), 48–52. doi:10.1145/2425248.2425259

Cook, G. (2014). Clicking Clean: How Companies are Creating the Green Internet. Retrieved from http://www.greenpeace.org/usa/Global/usa/planet3/PDFs/clickingclean.pdf

Falvo, M. C., Graditi, G., & Siano, P. (2014). Electric Vehicles Integration in Demand Response Programs. In *International Symposium on Power Electronics, Electrical Drives, Automation and Motion* (pp. 548–553). IEEE.

Ghatikar, G., Ganti, V., Matson, N., & Piette, M. A. (2012). Demand response opportunities and enabling technologies for data centers: Findings from field studies. Retrieved from https://escholarship.org/uc/item/7bh6n6kt

Goddard, G., Klose, J., & Backhaus, S. (2014). Model development and identification for fast demand response in commercial HVAC systems. *IEEE Transactions on Smart Grid*, *5*(4), 2084–2092. doi:10.1109/TSG.2014.2312430

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, *37*(2), 337–356. Retrieved from http://misq.org/positioning-and-presenting-design-science-research-for-maximum-impact.html

Heilig, L., & Voß, S. (2014). Decision Analytics for Cloud Computing: A Classification and Literature Review.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, *28*(1), 75–105.

IEA. (2014). Key World Energy Statistics 2014. Retrieved from https://www.iea.org/publications/freepublications/publication/key-world-energy-statistics-2014.html

Jheng, J. J., Tseng, F. H., Chao, H. C., & Chou, L. Der. (2014). A novel VM workload prediction using Grey Forecasting model in cloud data center. In *The International Conference on Information Networking 2014 (ICOIN2014)* (pp. 40–45). IEEE. doi:10.1109/ICOIN.2014.6799662

Kirschen, D. S. (2003). Demand-side view of electricity markets. *IEEE Transactions on Power Systems*, *18*(2), 520–527. doi:10.1109/TPWRS.2003.810692

Kishore, T. S., & Singal, S. K. (2014). Optimal economic planning of power transmission lines: A review. *Renewable and Sustainable Energy Reviews*, *39*, 949–974. doi:10.1016/j.rser.2014.07.125

Koomey, J. G. (2011). Growth in data center electricity use 2005 to 2010. Retrieved from http://www.analyticspress.com/datacenters.html

Lütticke, M. (2014). Opponents strike back at planned power line. Retrieved March 31, 2015, from http://www.dw.de/opponents-strike-back-at-planned-power-line/a-17410969

Mack, C. A. (2011). Fifty years of Moore's Law. *IEEE Transactions on Semiconductor Manufacturing*, *24*(2), 202–207. doi:10.1109/TSM.2010.2096437

Moore, G. E. (1965). Cramming more components onto integrated circuits. *Electronics*, *38*(8), 114–117. Retrieved from ftp://download.intel.com/sites/channel/museum/Moores_Law/Articles-Press_Releases/Gordon_Moore_1965_Article.pdf

NERC. (2011). NERC IVGTF Task 2.4 Report Operating Practices, Procedures, and Tools. Retrieved April 5, 2015, from http://www.nerc.com/files/ivgtf2-4.pdf

Rao, L., Liu, X., Xie, L., & Liu, W. (2012). Coordinated Energy Cost Management of Distributed Internet Data Centers in Smart Grid. *IEEE Transactions on Smart Grid*, *3*(1), 50–58. doi:10.1109/TSG.2011.2170100

regelleistung.net. (2015). Retrieved March 31, 2015, from https://www.regelleistung.net/ip/action/index

Stetz, T., von Appen, J., Niedermeyer, F., Scheibner, G., Sikora, R., & Braun, M. (2015). Twilight of the Grids. *Power and Energy Magazine, IEEE*, *13*(2), 50–61.

Van Hulle, F., Fichaux, N., Sinner, A. F., Morthost, P. E., Munksgaard, J., & Ray, S. (2010). Powering Europe: Wind Energy and the Electrical Grid. EWEA. Retrieved from http://www.ewea.org/fileadmin/files/library/publications/reports/Grids_Report_2010.pdf

Vandezande, L., Meeus, L., Belmans, R., Saguan, M., & Glachant, J.-M. (2010). Well-functioning balancing markets: A prerequisite for wind power integration. *Energy Policy*, *38*(7), 3146–3154. doi:10.1016/j.enpol.2009.07.034

Wang, H., Huang, J., Lin, X., & Mohsenian-Rad, H. (2014). Exploring smart grid and data center interactions for electric power load balancing. *ACM SIGMETRICS Performance Evaluation Review*, *41*(3), 89–94. doi:10.1145/2567529.2567556

Zhang, Y., Wang, Y., & Wang, X. (2011). Greenware: Greening cloud-scale data centers to maximize the use of renewable energy. In *Middleware 2011* (pp. 143–164). Springer Berlin Heidelberg.