

Online Electricity Cost Saving Algorithms for Co-Location Data Centers

Linquan Zhang[†], Zongpeng Li[†], Chuan Wu[‡], Shaolei Ren^{*}

[†]Department of Computer Science, University of Calgary

[‡]Department of Computer Science, The University of Hong Kong

^{*}School of Computing and Information Sciences, Florida International University
{linqzhan,zongpeng}@ucalgary.ca, cwu@cs.hku.hk, sren@fiu.edu

ABSTRACT

This work studies the online electricity cost minimization problem at a co-location data center. A co-location data center serves multiple tenants who rent the physical infrastructure within the data center to run their respective cloud computing services. Consequently, the co-location operator has no direct control over power consumption of its tenants, and an efficient mechanism is desired for eliciting desirable consumption patterns from the co-location tenants. Electricity billing faced by a data center is nowadays based on both the total volume consumed and the peak consumption rate. This leads to an interesting new combinatorial optimization structure on the electricity cost optimization problem, which also exhibits an online nature due to the definition of peak consumption. We model and solve the problem through two approaches: the *pricing* approach and the *auction* approach. For the former, we design an offline 2-approximation algorithm as well as an online algorithm with a small competitive ratio in most practical settings. For the latter, we design an efficient $(2 + c)$ -competitive online algorithm, where c is a system dependent parameter close to 1.49, and then convert it into an efficient mechanism that executes in an online fashion, runs in polynomial time, and guarantees truthful bidding and $(2 + 2c)$ -competitive in social cost.

Categories and Subject Descriptors

C.4 [Performance of Systems]: Design studies; Modeling techniques; I.1.2 [Algorithms]: Analysis of algorithms

Keywords

Co-Location Data Center; Mechanism Design; Approximation Algorithms; Online Algorithms

1. INTRODUCTION

Co-location data centers (or *co-locations*) rent physical space and infrastructure support, *e.g.*, reliable power supply and cooling service, to multiple tenants for hosting their servers at a common site. Co-locations have become an indispensable part for today's cloud computing industry, offering a flexible data center solution to

small and medium users who wish to run their own 'cloud' but are otherwise deterred by the daunting cost of constructing and maintaining their own data center. Even large users, as exemplified by Google and Akamai, rely on co-locations as a cost-effective complement to their own data centers for achieving a global presence, particularly in regions of relatively low demand that do not justify a dedicated data center.

Electricity charges paid by a co-location is computed by an interesting formula that includes two components: i) the *peak charge*, which is determined by the peak demand within a billing cycle, *e.g.*, the maximum average power consumption measured over each 15-minute interval; ii) the *volume charge*, which is based on total energy consumption in the billing cycle [1, 3]. The volume charge is relatively intuitive. The rationale behind the peak charge is that peak consumption shedding is critical to a power grid; even a small reduction in peak demand can provide significant cost savings, reduce greenhouse gas emission, and help balance grid-wide demand and supply. In practice, the peak charge component is seen to account for over 30% of the total electricity bill [3]. For example, consider a co-location data center located in British Columbia, Canada, powered by BC Hydro [1] with 24 MW peak demand and 15 MW average demand. The monthly peak charge is \$238,800, while the volume charge is \$524,880. In this case, the peak charge is 31% of the total monthly payment. This suggests that a *well designed algorithm for shaping the power consumption profile and controlling the peak demand has a great potential in helping cut electricity cost at a co-location.*

However, different from the case of private data centers, the co-location operator has no direct control on which machines are on/off, since its role is to offer basic services such as stable power supply and cooling. The individual tenants at a co-location manage their own servers and control the corresponding power consumption. While the co-location has a strong incentive to cut peak consumption and therefore save cost, its tenants may or may not share that same interest, depending on the contract between the two sides.

Typical electricity pricing today between a co-location and its tenants is flat-rate based, and does not depend on the real consumption volume or pattern [2, 4]. The tenants have little incentive to reduce their electricity usage by shutting down under-utilized servers, or by modulating their consumption pattern via shifting computing jobs in the temporal domain to reduce peak consumption rate. Such actions desired by the co-location will not automatically happen without appropriate incentives. Co-locations are sometimes so desperate to cut peak consumption that they start their own stand-by generators to cover part of the demand from its tenants [3]. Such quick-start stand-by generation (*e.g.*, using diesel generators) is often not economical, nor is it environment friendly. Auction based demand response mechanisms are natural approaches that have the

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).

SIGMETRICS'15, June 15-19, 2015, Portland, OR, USA.

ACM 978-1-4503-3486-0/15/06.

<http://dx.doi.org/10.1145/2745844.2745894>.

potential of efficiently providing incentive for tenants to cooperate, eliciting desired electricity consumption patterns with remuneration paid in return. A well designed auction may represent a win-win solution for both the co-location operator and its tenants, such that both enjoy a higher utility.

The maximum power demand is dependent on the power consumption in all time intervals during a billing cycle. Decisions in different time slots are therefore coupled, leading to an inherent online nature of the problem of demand shaping. Appropriate control of the peak demand helps reduce electricity cost while inappropriate control might incur a high cost. For a quick illustration, assume the energy demand pattern is $(D, D/2, D/2, D/2)$, possible savings by tenants are $(D/2, 0, 0, 0)$, the price for energy reduction is $f/2$, peak and volume charges are f and $f/4$, respectively. The optimal algorithm reduces the peak demand to $D/2$ at the beginning by using energy reduction from tenants, so that the total cost is $5Df/4$; while an online algorithm without future information might anticipate high demand in the future and refuse to use energy reduction from tenants, which makes the total cost $13Df/8$, 30% higher than the optimum. Even for the offline version of the problem (future prices and demands are perfectly known), computing the optimal solution efficiently is still highly non-trivial, since its underlying optimization problem is an integer program, which is NP-hard in general. The challenge escalates when one seeks to design an *online solution* for practical application, as knowledge on the demands, unit power prices as well as tenants' bids in the future are completely unknown.

2. RESULTS OVERVIEW

In this work, we model and solve the electricity cost saving problem in a co-location data center through two approaches: the *pricing* approach and the *auction* approach. In the pricing approach, the co-location data center offers a price it is willing to pay for unit energy reduction by tenants, and the tenants decide and submit how much energy they are willing to save at that price. Finally the co-location data center determines which tenants' energy reductions are accepted. In the auction approach, the co-location data center invites the tenants to submit energy reduction bids including the amount of energy consumption to shed and the amount of remuneration asked. The co-location then conducts a reverse auction to determine winning bids along with their corresponding payments. The pricing approach is relatively simple and has been applied in real world demand response solutions, implemented in electric appliances on the market. It is simple but requires the co-location to first come up with a good estimate on a unit reduction offer. The auction approach eliminates the need of such ad hoc guesses and resorts to the power of the market instead for automatic fair price revelation based on demand and supply. However, the auction design is considerably more complex than the algorithm design in the pricing approach, and our solution to the former borrows techniques from the latter.

Pricing Approach. We first study the offline cost minimization problem, find a piecewise linear function between the peak demand and the total electricity cost, and then propose a 2-approximation algorithm with future information assumed. By applying the approximation algorithm on the information currently available to decide the amount of energy obtained from the grid, and the primal-dual framework on the remaining linear integer program, we then proceed to design an online algorithm, which strives to increase the peak demand gracefully as requests and bids arrive. The online algorithm achieves a competitive ratio of $(1 + 2(\kappa + 1)/\rho + 2)$, where κ and ρ are system dependent parameters. The ratio is 6.2 when $\kappa = 3$ and $\rho \approx 2.5$ in our empirical studies. The ratio has two

parts: the term 2 results from addressing the computational complexity in the integer programming nature of the problem, while the term $1 + 2(\kappa + 1)/\rho$ results from addressing the challenge of the online nature of peak demand control.

Auction Approach. A reverse auction is conducted by the co-location data center. We model social cost minimization using a natural mixed integer program. We first analyze the social cost minimization problem by deriving the dual of its linear program relaxation. We design a $(2 + \xi/(1 - \frac{1}{\rho}))$ -competitive online algorithm, by applying the primal-dual framework to iteratively compute primal and dual solutions that are feasible to the primal and dual problems, respectively. The dual solution acts as a lower bound of the primal optimum, which helps analyze the competitive ratio. In empirical studies, we find $\xi \approx 1.2$ and $\rho \approx 2.5$, resulting in a competitive ratio of 4. Similar to the pricing approach, 2 comes from addressing challenges in computational complexity while $\xi/(1 - \frac{1}{\rho})$ comes from coping with lack of future information.

We study improvements to reduce the competitive ratio of the online algorithm, by keeping track of the maximum demand that has happened so far, and carefully exploiting that extra piece of information in decision making. That reduces the competitive ratio to $(2 + c)$, where c is a system dependent parameter. We find $c \approx 1.49$ in our empirical studies, and therefore the competitive ratio is 3.49, which is improved from 4. The new online algorithm solves two sub-problems: i) deciding the energy drawn from the grid based on historic information as well as current demand; ii) using the bids from tenants to fill the gap between current demand and the energy drawn from the grid. The second sub-problem is a general version of the integer program in the pricing approach, and a similar approximation algorithm is applied.

Since the tenants decide their bidding prices, a strategic tenant may submit a falsified price in the hope of a higher payment from the co-location. A *truthful* auction is desired, in which it is a dominant strategy for each tenant to bid its true cost. The celebrated Vickrey-Clarke-Groves (VCG) auction guarantees both truthfulness and social cost minimization. Unfortunately, a VCG auction requires solving the underlying optimization problem, which is NP-hard in our work, for multiple times, making it impractical. We resort to a randomized auction design framework, which decomposes a fractional optimal solution to the second sub-problem into a convex combination of feasible integer solutions. The framework runs in polynomial time, ensures the winning probability of a bid is monotonically non-increasing in its bidding price, and therefore is truthful in expectation. Combining the randomized auction framework with the online algorithm, we obtain an online auction with a competitive ratio of $(2 + 2c)$, which has only a small loss over that of the online algorithm.

Acknowledgments. This work has been supported by Wedge Networks, NSERC, AITF, HK RGC (HKU 718513), and NSF (CNS-1423137 and CNS-1453491).

3. REFERENCES

- [1] BC Hydro. <https://www.bchydro.com/>.
- [2] Enaxis Consulting. Pricing Data Center Co-location Services, 2009, <http://enaxisconsulting.com>.
- [3] Z. Liu, A. Wierman, Y. Chen, B. Razon, and N. Chen. Data Center Demand Response: Avoiding the Coincident Peak via Workload Shifting and Local Generation. In *Proc. of ACM SIGMETRICS*, 2013.
- [4] A. Qureshi, R. Weber, H. Balakrishnan, J. Gutttag, and B. Maggs. Cutting the Electric Bill for Internet-scale Systems. In *Proc. of ACM SIGCOMM*, 2009.