Algorithms for Dynamic Capacity Provisioning

Lachlan L. H. Andrew*, Minghong Lin†, Adam Wierman† and Zhenhua Liu†
*Centre for Advanced Internet Architectures, Swinburne University of Technology, Australia
†Dept. of Computer Science, California Institute of Technology, USA

***Abstract*—Data centre power consumption can be reduced by switching off servers during low load. However, excess switching is wasteful. This paper reviews online algorithms for optimizing this tradeoff, including the benefits of shifting load between geographically distant data centres. These algorithms can also adjust a link's number of parallel lightpaths.**

I. Introduction

The world is facing a need to curb its consumption of fossil fuels. The internet is an important tool for reducing this consumption by "dematerialization", which substitutes physical transport and goods by online equivalents. However, the data centres behind the internet are themselves major consumers of energy. This paper reviews recent results on algorithms for reducing this total energy consumption, or for reducing the amount of non-renewable energy consumed.

The techniques considered here all save energy by turning servers off during periods of low load. This is made feasible by the ability to migrate a virtual machine from one physical machine to another, but still incurs a substantial cost. The algorithms seek to balance this switching cost against the cost of operating excess servers. A central challenge is that the future load is unknown or only partially known. The algorithms provably guarantee that the cost is "close" to the minimum cost that could have been achieved had the future load been known.

II. Model

Consider a system of S data centres, each containing M servers. Work is sent to these data centres from different locations, indexed by j. At time t, source j requires that work $\lambda_{t,j}$ be performed, which may be split between the data centres. The costs incurred by data centre s at time t depend on the number of servers that are powered on, $x_{t,s}$, and the vector $\lambda_{t,s}$ of amounts of work it performs for the sources. The running cost $\mathcal{R}_{t,s}(x_{t,s}, \lambda_{t,s})$ accounts for server energy, queueing delays and network latency, and is jointly convex in all variables. There is also a switching cost β_s modelling the total cost of turning a server on and off, including the energy, delay, wear-and-tear and risk. We seek to solve

$$\begin{aligned} \min_{\substack{x_{t,s}\\ \lambda_{t,j,s}}} \quad & \sum_{t=1}^{T}\sum_{s=1}^{S} \mathcal{R}_{t,s}(x_{t,s},\lambda_{t,s}) + \beta_s(x_{t-1,s},x_{t,s})^+ \qquad (1)\\ \text{s.t.} \quad & \sum\nolimits_{s=1}^{S} \lambda_{t,j,s} = \lambda_{t,j}, \quad \forall t, \forall j \\ & \lambda_{t,j,s} \geq 0, \quad \forall t,\forall j,\forall s \\ & 0 = x_{0,s} \leq x_{t,s} \leq M, \quad \forall t, \forall s \end{aligned}$$

where $(x)^+ = \max(0, x)$. The challenge is that $x_{t,s}$ and $\lambda_{t,j,s}$ must be chosen by time t, without full knowledge of future loads or energy prices at times $\tau > t$.

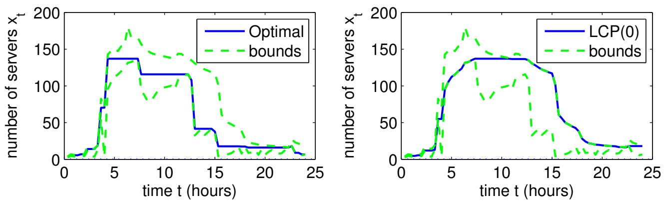

Fig. 1. Lazy capacity provisioning [1]. (a) Optimal solution, lazy backward in time. (b) LCP, lazy forward in time.

III. Algorithms

Efficient algorithms are known for several important special cases of (1). These algorithms do not use future knowledge, except for a small window of size $w \geq 0$. They are provably good in the sense that their performance is close to that of the optimal algorithm that has complete future knowledge.

Lazy Capacity Provisioning (LCP): If all servers are homogeneous, say within one data centre, then $S = 1$ and (1) becomes a one-dimensional optimization over x_t. In this case, the optimal solution has a very elegant structure, in terms of natural lower and upper bounds on the optimal value of x_t corresponding to the future load being either zero or very high. Starting from time $t = T$ and working backwards to $t = 1$, x remains between these bounds, and changes "lazily": only as much as is necessary not to violate the bounds, as shown in Figure 1(a).

LCP [1] is an implementable approximation to this structure, in which x remains within the same bounds, but is lazy forward in time, rather than backward, as in Figure 1(b). The bounds used by LCP can be calculated with no knowledge of future workloads, giving the scheme LCP(0). Tighter bounds can be obtained if information is available about a window w into the future, giving a family of schemes LCP(w).

LCP is also suitable for dynamically adapting the number of lightpaths to use on a particular optical path. It has often been proposed that power can be saved in optical links by turning off transcievers on some wavelengths sharing a fibre [2]. However, such studies often neglect the cost of turning wavelengths on and off. Turning a wavelength on or off causes transients in amplifiers along the path, which can cause high bit error rates in other lightpaths, jeopardizing service level agreements. This cost can be modelled by β, which allows LCP to trade off the switching cost against the cost of energy.

Receding horizon control: Daily patterns in data centre workloads can be predicted in advance. This allows the use of the classical approach of receding horizon control, RHC, (also called model predictive control, MPC). At time t, this calculates the optimal control over time t to $t + w$, based on predicted load, and implements that control for only the current timestep. In the next time step, it repeats the process using an updated prediction for the interval $t+1$ to $t+w+1$. If all servers are equivalent, such as within a single data centre, then this performs well for large windows; its maximum suboptimality decays as β/w [3]. Unfortunately, it can perform poorly when load is to be distributed over many types of servers, as is the case with geographical load balancing. However large w is, there exist bad cases where the cost can be up to $(1+\beta)$ times the optimal cost, which is as bad as the case with no load prediction [3].

Averaging fixed horizon control: When there are multiple data centre case, the weakness of RHC can be overcome by using a new technique proposed in [3]. This technique makes use of old predictions as well as new ones. Specifically, at time t it implements the *average* of the controllers designed in each of the past w time steps, instead of simply the most recent controller as used by RHC. This performs similarly to RHC in typical cases, but the maximum suboptimality is guaranteed to decay as β/w, regardless of how many types of server there are.

The missing case: The three foregoing results can be classified by whether or not there are multiple data centres, and whether or not there is workload prediction. The remaining case is when there are multiple data centres but no prediction. There is no known competitive algorithm for this case in general. The problem is a special case of a metrical task system (MTS) but direct use of the standard MTS algorithms results performance that degrades as the number of servers increases. However, recent results suggest that a solution may be available in the special case that $\mathcal{R}_{t,s}(x,\lambda)$ is linear in x when $x \geq \lambda$ and infinite otherwise. In this case, applying an MTS algorithm separately to carefully chosen subsets of the workload results in an algorithm whose performance depends only on the (small) number of data centres, rather than the (much larger) number of servers.

IV. Numerical results

Geographical load balancing has been proposed to reduce energy costs [4]. Paradoxically, this increases the total energy consumption; by reducing the average cost per kWh, it makes high energy use more affordable. However, geographical load balancing provides a powerful tool for exploiting renewable energy [5], [6].

The biggest challenge to relying on renewable energy is that sources such as solar and wind are intermittent. Since energy storage is expensive and transmission is subject to capacity constraints, supply must approximately match demand at all times and in all regions. If supply cannot be controlled, then demand must be controlled, using "demand response" mechanisms [7]. The ability to shift load between data centres allows energy demand to be shifted to places where renewable energy is currently available. However, this only occurs if the price of electricity is tied to the current availability.

This is illustrated in Figure 2, taken from [5], which considers the number of active servers required in a model of Google's US data centres. Energy is a mix of solar, subject to daily fluctuations, and controllable non-renewable energy. In the first example, the price of energy is fixed, and the number of servers follows an expected diurnal pattern. In the second, renewable energy is priced at one tenth non-renewable energy, to reflect its lower *incremental* cost. This causes evening load on the east cost to be routed to the west coast where solar energy is still available, and morning load to be routed east.

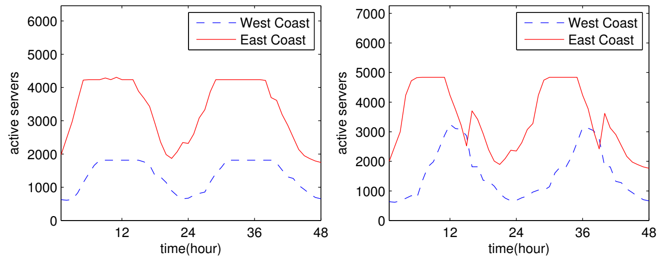

Fig. 2. Number of active servers under static pricing and when solar energy is abundant

V. Conclusion

Along with the engineering improvements to data centre energy efficiency, there is a need for algorithms to manage the infrastructure. The algorithms mentioned above provide a set of tools for such management.

Geographical load balancing's benefits bring drawbacks. First, the peak number of servers required in the second scenario is noticeably higher in the second scenario than the first, suggesting an increase in both the embodied energy of equipment and its financial cost. Second, the amount of data being transported is much higher in the second scenario, which increases the network energy consumption. Finally, it requires agile routing, capable of transferring load on a relatively fast time scale without interrupting current user sessions. These are important ongoing research directions.

VI. Acknowledgements

This work was supported by NSF grants CCF 0830511, CNS 0911041, and CNS 0846025, MURI grant W911NF-08-1-0233, Microsoft Research, the Lee Center for Advanced Networking, and Australian Research Council grant FT0991594.

References

[1] M. Lin, A. Wierman, L. L. H. Andrew, and E. Thereska, "Dynamic right-sizing for power-proportional data centers," in *Proc. INFOCOM*, 2011.

[2] W. Fisher, M. Suchara, and J. Rexford, "Greening backbone networks: reducing energy consumption by shutting off cables in bundled links," in *Proc. ACM worksh. Green networking*. ACM, 2010, pp. 29–34.

[3] M. Lin, Z. Liu, A. Wierman, and L. L. H. Andrew, "Online algorithms for geographical load balancing," in *Proc. Int. Green Computing Conf.*, San Jose, CA, Jun. 2012.

[4] A. Qureshi, R. Weber, H. Balakrishnan, J. Guttag, and B. Maggs, "Cutting the electric bill for internet-scale systems," in *ACM SIGCOMM*, Aug. 2009, pp. 123–134.

[5] Z. Liu, M. Lin, A. Wierman, S. H. Low, and L. L. H. Andrew, "Greening geographical load balancing," in *Proc. ACM SIGMETRICS*, San Jose, CA, 7-11 Jun 2011, pp. 233–244.

[6] ——, "Geographic load balancing with renewables," in *Proc. ACM GreenMetrics*, San Jose, CA, 7 Jun 2011.

[7] M. Albadi and E. El-Saadany, "Demand response in electricity markets: An overview," in *IEEE Power Engineering Society General Meeting*. IEEE, 2007, pp. 1–5.