Temperature Aware Workload Management in Geo-distributed Datacenters

Hong Xu, Chen Feng, Baochun Li
Department of Electrical and Computer Engineering
University of Toronto
{henryxu, cfeng, bli}@eecg.toronto.edu

ABSTRACT

Datacenters consume an enormous amount of energy with significant financial and environmental costs. For geo-distributed datacenters, a workload management approach that routes user requests to locations with cheaper and cleaner electricity has been shown to be promising lately. We consider two key aspects that have not been explored in this approach. First, through empirical studies, we find that the energy efficiency of the cooling system depends directly on the ambient temperature, which exhibits a significant degree of geographical diversity. Temperature diversity can be used by workload management to reduce the overall cooling energy overhead. Second, energy consumption comes from not only interactive workloads driven by user requests, but also delay tolerant batch workloads that run at the back-end. The elastic nature of batch workloads can be exploited to further reduce the energy cost.

In this work, we propose to make workload management for geo-distributed datacenters *temperature aware*. We formulate the problem as a joint optimization of request routing for interactive workloads and capacity allocation for batch workloads. We develop a distributed algorithm based on an m-block *alternating direction method of multipliers* (ADMM) algorithm that extends the classical 2-block algorithm. We prove the convergence and rate of convergence results under general assumptions. Trace-driven simulations demonstrate that our approach is able to provide 5%–20% overall cost savings for geo-distributed datacenters.

Categories and Subject Descriptors

C.2.4 [**Computer-Communication Networks**]: Distributed Systems

Keywords

Geo-distributed datacenters; Energy; Request routing; Cooling efficiency; ADMM; Distributed optimization

1. INTRODUCTION

Geo-distributed datacenters operated by organizations such as Google and Amazon are the powerhouses behind many Internet-scale services. They are deployed across the globe to provide better latency and redundancy. These datacenters run hundreds of thousands of servers, consume megawatts of power with massive carbon footprint, and incur electricity bills of millions of dollars [2, 6]. Recently, important progress has been made on a new *workload management* approach that focuses on the energy aspect of geo-distributed datacenters. It exploits the geographical diversity of electricity prices by optimizing the *request routing* algorithm to route user requests to locations with cheaper and cleaner electricity [2, 5–7].

In this work, we consider two key aspects of geo-distributed datacenters that have not been explored in the existing literature.

First, cooling systems, which consume 30% to 50% of the total energy, are often modeled with a constant and location-independent energy efficiency factor in existing efforts. This tends to be an oversimplification in reality. Through our study of a state-of-the-art production cooling system, we find that temperature has direct and profound impact on cooling energy efficiency. This is especially true with *outside air cooling* technology, which has seen increasing adoption in mission-critical datacenters [8]. As shown in Table 1, the partial PUE (power usage effectiveness), defined as the sum of server power and cooling overhead divided by server power, varies significantly from 1.30 to 1.05 when temperature drops from 35 °C(90 °F) to -3.9 °C(25 °F). The reason is that when the ambient temperature is low, we can directly use the cold outside air to cool down servers without running the energy-gobbling mechanical chillers, which greatly improves the energy efficiency.

Outdoor ambient	Cooling mode	pPUE
35°C(90°F)	Mechanical	1.30
21.1°C(70°F)	Mechanical	1.21
15.6°C(60°F)	Mixed	1.17
10°C(50°F)	Outside air	1.1
-3.9°C(25°F)	Outside air	1.05

Table 1: Efficiency of Emerson's DSE™ cooling system with an EconoPhase air-side economizer [8]. Return air is set at 29.4°C(85°F).

Through an extensive empirical analysis of daily and hourly climate data for 13 Google datacenters [8], we further find that temperature varies significantly across both time and location, which is intuitive to understand. The short-term volatilities are not well correlated across locations. These observations suggest that datacenters at different locations have distinct and time-varying cooling energy efficiency. This establishes a strong case for making workload management *temperature aware*, where such temperature diversity can be used along with price diversity in making request routing decisions to reduce the overall cooling energy overhead.

Second, energy consumption comes not only from interactive workloads driven by user requests, but also from delay tolerant batch workloads, such as indexing and data mining jobs, that run at the back-end. Existing efforts focus mainly on request routing to minimize the energy cost of interactive workloads, which is only a part of the entire picture. Such a mixed nature of datacenter workloads provides more opportunities to utilize the cost diversity of

SIGMETRICS'13, June 17–21, 2013, Pittsburgh, PA, USA.
ACM 978-1-4503-1900-3/13/06.

energy. The key observation is that batch workloads are elastic to resource allocations, whereas interactive workloads are highly sensitive to latency and have more profound impact on revenue. Thus at times when one location is comparatively cost efficient, we can increase the capacity for interactive workloads by reducing the resources reserved for batch jobs. More requests can then be routed to and processed at this location, and the cost saving can be more substantial. We are thus motivated to advocate a holistic workload management approach, where *capacity allocation* between interactive and batch workloads is dynamically optimized with request routing.

2. CONTRIBUTIONS

Towards temperature aware workload management, we propose a general framework to capture the important trade-offs involved [8]. We model both energy cost and utility loss, which corresponds to performance-related revenue reduction. We develop an empirical cooling efficiency model based on the production system in Table 1 with both outside air and mechanical cooling capabilities. The problem is formulated as a joint optimization of request routing and capacity allocation. The technical challenge is then to develop a distributed algorithm to solve the large-scale optimization with tens of millions of variables for a production geo-distributed cloud. Dual decomposition with subgradient methods is often used to develop distributed optimization algorithms. However it requires delicate adjustment of step sizes that makes convergence difficult to achieve for large-scale problems. The method of multipliers achieves fast convergence, at the cost of introducing tight coupling among variables.

We rely on the *alternating direction method of multipliers* (ADMM), a simple yet powerful algorithm that blends the advantages of the two approaches. ADMM recently has found practical use in many large-scale distributed convex optimization problems [1]. It works for problems whose objective and variables can be divided into *two* disjoint parts. It alternatively optimizes part of the objective with one block of variables to iteratively reach the optimum. Our formulation has three blocks of variables, yet little is known about the convergence of m-block ($m \geq 3$) ADMM algorithms, with two exceptions [3, 4] very recently. [3] establishes the convergence of m-block ADMM for strongly convex objective functions, but not linear convergence; [4] shows the linear convergence of m-block ADMM under the assumption that the relation matrix is full column rank, which is, however, not the case in our formation. This motivates us to refine the framework in [4] so that it can be applied to our setup. In particular, we show that by replacing the full-rank assumption with some mild assumptions on the objective functions, we are not only able to obtain the same convergence and rate of convergence results, but also to simplify the proof of [4] in [8]. The m-block ADMM algorithm is general and can be applied in other problem domains. For our case, we further develop a distributed algorithm, which is amenable to a parallel implementation in datacenters.

3. EVALUATION

We conduct extensive trace-driven simulations with Wikipedia request traffic traces, real-world electricity prices and historical temperature data at Google datacenter locations to realistically assess the potential of our approach [8]. We benchmark our ADMM algorithm against the state-of-the-art approach, which is a temperature agnostic strategy that separately considers capacity allocation and request routing of the workload management problem. It first allocates capacity to batch jobs by minimizing the back-end total cost as the objective. The remaining capacity is used to solve the request routing optimization. Only the electricity price diversity is used, and cooling energy is calculated with a constant pPUE of 1.2 for the two cost minimization problems. Though naive, such an approach is widely used in current Internet-scale cloud services. It also allows for an implicit comparison with prior work [2, 5–7]. We run the algorithms with our 24-hour traces at each day of January, May, and August 2011 [8]. The results are averaged over 31 runs.

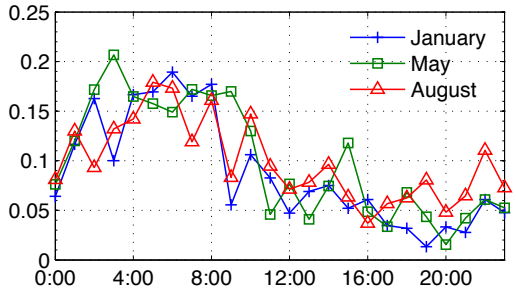

Figure 1: Overall cost savings of our approach compared to state-of-the-art workload management.

Figure 1 shows the average overall cost savings, including energy cost savings and utility loss reductions. We observe that the cost savings range from 5% to 20%. This shows that our approach is able to provide substantial cost savings for geo-distributed datacenters, using temperature-aware request routing and dynamic capacity allocation. The savings are also consistent and insensitive to seasonal changes. The reason is that our approach depends on: 1) the geographical diversity of temperature and cooling efficiency; 2) the mixed nature of datacenter workloads, both of which exist at all times of the year no matter which cooling method is used. Temperature aware workload management is thus expected to offer consistent and promising cost benefits.

References

[1] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2010.

[2] P. X. Gao, A. R. Curtis, B. Wong, and S. Keshav. It's not easy being green. In *Proc. ACM SIGCOMM*, 2012.

[3] D. Han and X. Yuan. A note on the alternating direction method of multipliers. *J. Optim. Theory Appl.*, 155:227–238, 2012.

[4] M. Hong and Z.-Q. Luo. On the linear convergence of the alternating direction method of multipliers. `http://arxiv.org/abs/1208.3922`, August 2012.

[5] Z. Liu, M. Lin, A. Wierman, S. H. Low, and L. L. Andrew. Greening geographical load balancing. In *Proc. ACM Sigmetrics*, 2011.

[6] A. Qureshi, R. Weber, H. Balakrishnan, J. Guttag, and B. Maggs. Cutting the electricity bill for Internet-scale systems. In *Proc. ACM SIGCOMM*, 2009.

[7] L. Rao, X. Liu, L. Xie, and W. Liu. Minimizing electricity cost: Optimization of distributed Internet data centers in a multi-electricity-market environment. In *Proc. IEEE INFOCOM*, 2010.

[8] H. Xu, C. Feng, and B. Li. Temperature aware workload management in geo-distributed datacenters. Technical report, University of Toronto, `http://iqua.ece.toronto.edu/~henryxu/share/geodc-preprint.pdf`, 2013.