Towards Cooling Internet-Scale Distributed Networks on the Cheap

Vani Gupta†, Stephen Lee†, Prashant Shenoy†, Ramesh K. Sitaraman†‡, Rahul Urgaonkar¶
†University of Massachusetts, Amherst ‡Akamai Technologies ¶IBM Research
{vani,stephenlee,shenoy,ramesh}@cs.umass.edu rurgaon@us.ibm.com

ABSTRACT

Internet-scale Distributed Networks (IDNs) are large distributed systems that comprise hundreds of thousands of servers located around the world. IDNs consume significant amounts of energy to power their deployed server infrastructure, and nearly as much energy to cool that infrastructure. We study the potential benefits of using renewable open air cooling (OAC) in an IDN. Our results show that by using OAC, a global IDN can extract 51% cooling energy reducing during summers and a 92% reduction in the winter.

Categories and Subject Descriptors

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

Keywords

Internet-Scale Distributed Systems, Load Balancing, Energy

1. INTRODUCTION

Internet-scale distributed Networks (IDNs) deliver web content, streaming media to clients via thousands of servers located in data centers throughout the world. These data centers consume significant amount of energy to power their servers and cool them. While the energy bills incurred in these data centers may run into millions of dollars per month, the environmental impact of IDNs is also substantial. With carbon emissions from data centers growing by 15% every year, it is imperative to re-design IDNs with energy as a key design consideration in order to ensure the sustainable growth of these networks. Since the energy bill spent in cooling an IDN is nearly as large as the portion used to power its servers, we focus explicitly on the problem of reduction in cooling energy usage using *open air cooling* (OAC) to reduce costs.

To our knowledge, OAC has not been studied in an IDN context and is distinctive for the following reasons. IDNs such as CDNs have two defining characteristics: (i) *global deployment* of servers in multiple data centers around the world, and (ii) *replication* of services across these data centers. The global deployment is often driven by the need for an IDN to have servers "proximal" to the end-users. Hence, it is not possible for IDNs to deploy data centers *only* in places where weather is cold most of the year, or where electricity is cheap. However, IDNs often replicate their services across their data centers, so that the workload of serving users can be easily shifted from one data center to another, albeit with a potential for performance degradation. A combination of these two characteristics provide an IDN the flexibility to move its workload across data centers to exploit climatic variations to optimize the use of OAC, a flexibility that services employing a single or a few data centers do not possess. Our algorithms orchestrate both the movement of load (i.e, load balancing) as well as energy drawn from the grid (i.e., power management) to decrease an IDN's cooling costs.

SIGMETRICS'15, June 15 -19, 2015, Portland, OR, USA.
ACM 978-1-4503-3486-0/15/06.
http://dx.doi.org/10.1145/2745844.2745898.

2. BACKGROUND

Internet-scale Distributed Networks: A recent study of data center energy consumption showed that servers and cooling consumed 56% and 30% of the total energy respectively. Since cooling energy is a significant portion of the total energy consumption, we examine the potential for employing OAC to reduce the energy usage in an IDN. We assume that an IDN (or, its proxy) has the ability to deploy or exploit advanced cooling technologies in all of its data centers locations. With increasing trend towards modular self-contained containers, it is reasonable to assume that IDNs will have greater control over how a data center is built and cooled.

Renewable Cooling using Open Air: A promising approach to reduce the cooling energy of a data center is to use the outside air, instead of chillers, to cool servers within the data center. However, OAC is feasible only when the air is "cold and dry enough". As a result, OAC may not be possible everywhere and may not be possible during all times of the day. In our work, we use (i) the instantaneous weather outside each of the IDN's data center locations, and (ii) the recommendations in Table 1 from the American Society of Heating, Refrigeration, and Air-conditioning Engineers (ASHRAE) to determine whether the outside air can be used to cool the data center at any point in time. To determine whether a data center can employ OAC at any point in time, we follow the methodology presented in the GreenGrid consortium whitepaper[3].

Data center Class	Dry-Bulb Temp (° C)	Humidity Range	Max Dew Point (° C)
A1	15 to 32	20% to 80%	17
A2	10 to 35	20% to 80%	21
A3	5 to 40	8% to 85%	24
A4	5 to 45	8% to 90%	24

Table 1: ASHRAE's allowable ranges suitable for cooling [3].

Energy Consumption Model: We use the standard linear model of server power consumption where the power (in Watts) consumed by a server serving load λ equals $P_{idle} + (P_{peak} - P_{idle})\lambda$, where

P_{idle} is the power consumed by an idle server, P_{peak} is the power consumed by the server under peak load, and λ is the load served by the server as a fraction of its capacity. Clearly, cooling energy is roughly proportional to the amount of heat dissipated in the data center, which in turn is governed largely by server energy. The proportionality factor is a measure of the effectiveness of the cooling system deployed at the data center and is captured by its power usage effectiveness (PUE) that is ratio of the total energy consumed by the facility to the energy consumed by the servers and other IT equipment. In our evaluation, we assume the average PUE is 1.80

3. EVALUATION

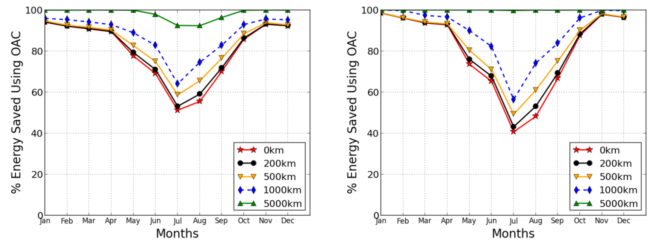
To evaluate the benefit of OAC, we used extensive load data from across Akamai's CDN for the period of one month, and global weather traces provided by the National Oceanic and Atmospheric Administration (NOAA) for the year 2012. We use a simple greedy algorithm that modifies the load assignments made by the (non-weather-aware) load balancer as reflected in our Akamai load traces by moving user load from data centers that have no OAC to nearby data centers that do, subject to performance constraints. Specifically, our greedy algorithm does the following for each of the IDN's data centers at each time step: if the weather conditions at a data center location permits OAC, then user load mapped to that location is unchanged; however, if the weather conditions at a data center location do not permit OAC to be employed, the load balancer attempts to greedily re-assign the load destined for that location to other nearby data centers (in increasing order of distance) with spare server capacity where OAC may be available. After our algorithm moves as much of the load as it can to open-air coolable data centers, the residual load must be cooled using chillers. Each unit of load that is cooled using OAC is a direct reduction in energy usage for the IDN, since chillers are not used. The energy savings can be computed by comparing the energy used with OAC to the energy used to cool the original load *entirely* with chillers.

3.1 Empirical Results

We evaluate the potential for OAC using our greedy algorithm outlined above on our IDN load and weather traces for a full year.

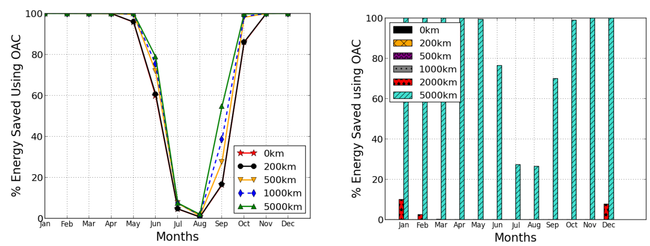
Reduction in global energy usage: Figure 1 depicts the average energy savings obtained across the entire CDN for different months of the year and for different values of distance r. The savings from OAC is higher during the cooler winter, with lower savings during the warmer summer months (May to September). Note that our analysis includes savings from data centers in both the hemispheres. However, Internet traffic from northern hemisphere dominate the global Internet traffic, hence the seasonal benefits from the northern hemisphere dominate the global trends. Overall, our result shows that even during summer, a global IDN can extract significant cooling energy reduction of more than 51% *with no performance impact* ($r = 0$) and the savings due to OAC increase to over 92% during winter months; when $r = 1000$km, an additional 13% savings during the warmest month of July. For $r = 5000$km, (allows trans-continental load redirection), the savings increase to nearly 100% throughout the year. We observed that the savings in USA broadly follow the global trends, indicating that USA not only contributes a significant portion of the global traffic, but also has its seasons aligned with the dominant northern hemisphere.

Regional and Seasonal Variations in Savings: Our results show significant regional and seasonal differences in the savings obtained from OAC. Japan (Figure 2(a)) has the most significant seasonal variations in OAC benefits—with nearly 100% energy savings from employing OAC during the winter months to a zero savings in the summer month of August for the no load redirection



(a) Global (b) USA

Figure 1: Energy savings for the entire global IDN



(a) Japan (b) Singapore

Figure 2: Regional and seasonal variations in OAC savings

scenario of $r = 0$. The savings do not increase significantly even when load can be redirected to other data centers with a radius of $r = 1000$km. While Japan exhibits the extreme seasonal variations in OAC benefits, Singapore (Figure 2(b)) sees the worst year-round benefits. While Singapore is itself a small country, it is a key regional "hub" with significant traffic. Since Singapore is located near the equator, it has warm and humid weather throughout the year with nearly no seasons. Consequently OAC is not possible in Singapore during any month of the year, yielding zero savings. Even allowing Singapore traffic to be sent to data centers within a 1000km radius yields no benefits.

4. CONCLUSIONS

In this poster we presented the potential benefits of using open air cooling to reduce the energy usage incurred by an IDN. We empirically evaluated the efficacy of OAC using extensive traces from Akamai's global CDN and global weather data from NOAA. Our results showed that, by using free OAC, a global IDN can extract 51% cooling energy reducing during summers and a 92% reduction in the winter.

5. ACKNOWLEDGEMENTS

This research is supported in part by National Science Foundation grants 1422245, 1229059 and 1413998.

6. REFERENCES

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

[2] Asfandyar Qureshi, Rick Weber, Hari Balakrishnan, John Guttag, and Bruce Maggs. Cutting the Electric Bill for Internet-Scale Systems. In *Proceedings of SIGCOMM, 2009*.

[3] Michael Patterson Tom Harvey and John Bean. Updated air-side free cooling maps: The impact of ASHRAE 2011 allowable ranges. *The Green Grid*, 2012.