MantisMail: Green Content Delivery for Email (Extended Abstract)

Nan Deng, Chen Li, and Christopher Stewart
Computer Science and Engineering
The Ohio State University

Xiaorui Wang
Electrical and Computer Engineering
The Ohio State University

Fiber optics have narrowed the gap between an ISP's total web traffic and bandwidth provided by high capacity links. Now, datacenters connected to ISPs via fiber links can quickly deliver web content to the ISP's users without placing edge servers within the ISP. Due to economies of scale, these datacenters can host web content efficiently, making it cheaper to set up a content delivery network (CDN) that caches content to reduce delivery times. Increasingly, CDNs must do more than reduce response times to attract customers; They also have to keep prices low and invest in new features that add value to their service.

We propose a new value-added feature for CDNs: Carbon offsetting. Carbon offsets represent a unit of work that voids 1g of carbon emission. By buying carbon offsets, CDNs can void emissions caused by traffic that they intercept, including emissions spent getting content from source servers. These *green CDNs* would add value because the growing carbon footprint of the web is concerning. If datacenters worldwide were treated as a country, they would rank among the top 30 carbon emitters. Routing, storing, and serving email alone causes more than $1M$ metric tonnes of carbon emission each year [9, 14], even though email is less than 1% of Internet traffic.

Research Challenge 1: To offset the emissions caused by retrieving content from source servers, green CDNs must estimate the carbon needs of servers and network links they do not control. Prior studies have estimated carbon needs for specific workloads, but most of these studies confound the contributions of applications, hardware, and environment. Our approach captures each independently, allowing us to predict the effects of Moore's Law and growing renewable energy penetration.

Research Challenge 2: The simplest design for a green CDN is 1) cache content on a cluster of servers, 2) size the cluster based on hit rate and bandwidth costs, 3) estimate emissions, and 4) buy offsets as needed. But carbon offsetting increases costs, affecting the best cluster size for a green CDN. Further, these costs vary over time with energy prices and renewable energy production.

$$\text{Min: } \lambda P_{c02e}[E_{source} \times M(N)] + \lambda P_B + P_{srvr} \times N \quad (1)$$

$$\text{Subject to: } \mu N \geq \lambda \quad (2)$$

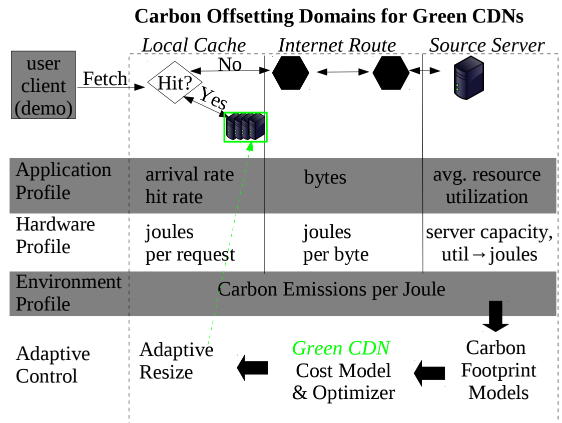

Figure 1: Data flow, profiling, and modeling in MantisMail—a green CDN. *At NSDI, we will demo MantisMail on our email accounts and allow attendees to test their own.*

Equations 1 & 2 preview our cost model that helps green CDN managers size their local cluster. N is the CDN cluster size. λ is the arrival rate of user requests. μ is the processing rate per instance; $M(N)$ is the miss rate function. E_{source} is the carbon footprint for getting content from the source. P_{c02e},$P_B$andP_{srvr} are the price of carbon offsets, outgoing bandwidth, and CDN servers. Our goal is to minimize cost with low response times.

MantisMail: Figure 1 depicts data flow, adaptive control, and carbon footprint modeling in *MantisMail*, our proof-of-concept green CDN. End users sign up for MantisMail by registering their email ID and source IMAP server address. Within their email clients, users configure MantisMail as their IMAP server. MantisMail caches entire IMAP requests using the Redis key value store, serving cached content when possible. Requests that can't be cached are sent to the user's source IMAP server. Note, MantisMail supports SSL and does not store passwords locally. Cached emails are encrypted.

Recently, the authors began accessing their email through MantisMail, offsetting 295g C02e per day (about 1 mile in an econ. car). While email is personal, cache hits occurred for users with multiple devices (12% hits). A content based cache could increase hits.

1 Extended Related Work

MantisMail builds on recent work in low carbon computing. In this extension, we discuss work in scheduling, data placement, and whole system design.

Scheduling Green Energy: Le et al. [10] studied request routing for Internet services that use multiple datacenters. Request routing is challenging because naive policies that reduce carbon footprints may lead to high response times. Le et al. [10] used simulated annealing and simulation to find sweet spots that lowered carbon footprints without increasing response times too much. Subsequent efforts formally solved the scheduling problem using constraint optimization [12] and demonstrated online approaches [11]. Zhang et al. [17] considered a different problem; Rather minimizing cost within a carbon constraint, they maximized the use of green energy within response time constraints. Others looked to extend this model to demand response scenarios for datacenters that work in collaboration with utility companies [2, 13].

Data Placement: Blink [15] proposed an API for cutting power from datacenters in response to lower renewable energy production. Blink also carefully placed data in key value stores to avoid performance degradation when servers are shut down. GreenHadoop adopts a similar strategy and couples a renewable energy prediction policy [7].

System Design: At the same time, many researchers have studied the design of real systems that are powered by clean energy. MantisMail is most like this group. MantisMail builds on the principle of green hosting [5] where clean energy is abstracted as carbon offsets that can be purchased from free markets. The challenge is to buy as few offsets as possible while maximizing revenue. Already many green hosting services has shown that investing in clean energy can increase profits [8].

Other work has looked into the design of datacenters with on-site renewable energy [1, 3, 4, 6, 16]. The focus of these efforts has been reducing costs.

References

[1] Parasol and greenswitch: Managing datacenters powered by renewable energy. In *Conference on Architectural Support for Programming Languages and Operating Systems*, 2013.

[2] D. Chiu, C. Stewart, and B. McManus. Electric grid balancing through low-cost workload migration. In *ACM GreenMetrics Workshop*, June 2012.

[3] N. Deng and C. Stewart. Concentrating renewable energy in grid-tied datacenters. In *International Symposium on Sustainable Systems and Technology (ISSST)*, May 2011.

[4] N. Deng, C. Stewart, D. Gmach, and M. Arlitt. Policy and mechanism for carbon-aware cloud applications. In *NOMS*, Apr. 2012.

[5] N. Deng, C. Stewart, J. Kelley, D. Gmach, and M. Arlitt. Adaptive green hosting. In *International Conference on Autonomic Computing*, 2012.

[6] D. Gmach, J. Rolia, C. Bash, Y. Chen, T. Christian, A. Shah, R. Sharma, and Z. Wang. Capacity planning and power management to exploit sustainable energy. In *International Conference on Network and Services Management*.

[7] I. n. Goiri, K. Le, T. D. Nguyen, J. Guitart, J. Torres, and R. Bianchini. Greenhadoop: leveraging green energy in data-processing frameworks. In *EuroSys Conf.*, 2012.

[8] Hostgator: Web hosting services. http://hostgator.com.

[9] ICF International. The carbon footprint of spam report, 2009.

[10] K. Le, O. Bilgir, R. Bianchini, M. Martonosi, and T. D. Nguyen. Capping the brown energy consumption of internet services at low cost. In *International Green Computing Conference*, Aug. 2010.

[11] M. Lin, Z. Liu, A. Wierman, and L. Andrew. Online algorithms for geographical load balancing. In *IEEE International Conference on Green Computing*, 2012.

[12] Z. Liu, M. Lin, A. Wierman, S. Low, and L. Andrew. Geographical load balancing with renewables. In *GreenMetrics Workshop*, June 2011.

[13] Z. Liu, A. Wierman, Y. Chen, and B. Razon. Data center demand response: Avoiding the coincident peak via workload shifting and local generation. In *poster at Proceedings of ACM Sigmetrics*, June 2013.

[14] S. Radicati. Email statistics report:executive summary, 2010.

[15] N. Sharma, S. Barker, D. Irwin, and P. Shenoy. Blink: managing server clusters on intermittent power. In *Conference on Architectural Support for Programming Languages and Operating Systems*, Mar. 2011.

[16] R. Singh, D. Irwin, P. Shenoy, and K. Ramakrishnan. Yank: Enabling green data centers to pull the plug. In *USENIX Symp. on Networked Systems Design and Implementation*, 2013.

[17] Y. Zhang, Y. Wang, and X. Wang. Greenware: Greening cloud-scale data centers to maximize the use of renewable energy. In *Middleware*, Dec. 2011.