Power Efficiency in Energy-aware Data Center Network

Tosmate Cheocherngngarn[1] [Jean H. Andrian[2] Deng Pan[3] Khokiat Kengskool[4]]

Abstract: Energy efficiency is becoming increasingly important in the operation of networking infrastructure, especially in enterprise and data center networks. Networking devices today consume a non-trivial amount of energy and it has been shown that this energy consumption is largely independent of the load through the devices. Numerous studies have shown that data center network rarely operate at full utilization, leading to a number of proposals for creating servers that are energy proportional with respect to the computation that they are performing. In this paper, as servers themselves become more energy proportional, the data center network can become a significant fraction of cluster power. We propose several ways to customize an energy-aware data center network whose power consumption is more proportional to the amount of traffic it is moving. Specifically, our approach is to propose cross-layer design for data center network.

Keywords: Energy Proportional; Energy-aware Data Center Network; Cross-layer Design

Introduction

Energy efficiency has become crucial for all industries, including the information technology (IT) industry, as there is a strong motivation to lower capital and recurring costs. With the advent of the Cloud Computing model, large data centers are being built that consolidate processing and storage for a large number of services accessed over the Internet or enterprise networks. In such environments, the initial focus on energy efficiency has been on cooling and server power management [1]–[2] and significant advances have been made in these areas. Recent studies have shown that networking devices account for about 15% of a data center's total energy consumption [3]. So far, scant attention has been paid to make networking in enterprise and data center networks more energy efficient. In this paper, we focus on energy savings algorithms for networking components in enterprise and data center networks that typically are under the control of a single administrative authority and thus making it possible to apply network-wide energy saving schemes.

Data centers have become popular computing infrastructure, because they achieve economies of scale with hundreds of thousands of servers, e.g. about 300,000 servers in Microsoft's Chicago data center [4]. At the same time, the huge number of servers in data centers consume significant amounts of energy. It is estimated that national energy consumption by data centers in 2011 will be more than 100 billion kWh, representing a $7.4 billion annual electricity cost. As a result, energy efficiency of data centers has attracted wide attention in recent years, mostly focusing on servers and cooling systems [2].

Ideally for power-efficiency, devices should consume energy proportional to their load [5]. The majority of the network devices deployed today is far from being energy proportional and provides a very limited set of knobs to

[1] Electrical and Computer Engineering Department, FIU College of Engineering and Computing 10555 West Flagler St. EC3900 Miami, FL 33174, Tosmate.Cheocherngngarn@fiu.edu

[2] Electrical and Computer Engineering Department, FIU College of Engineering and Computing 10555 West Flagler St. EC3900 Miami, FL 33174, Jean.Andrian@fiu.edu

[3] School of Computing and Information Sciences, FIU College of Engineering and Computing 10555 West Flagler St. EC3900 Miami, FL 33174, pand@fiu.edu

[4] Civil and Environmental Engineering Department, FIU College of Engineering and Computing 10555 West Flagler St. EC3900 Miami, FL 33174, kengskoo@fiu.edu

control their power consumption [6]. In this paper, we focus on how to achieve energy efficiency from these non-energy proportional devices. We propose several energy saving algorithms for efficient configuration and management of data center networks. We simulate the effects of these algorithms on Java Programing from an operational data center.

Servers	Cooling	Networking	Power Cond.	Lighting
36%	30%	25%	8%	1%

Table 1: Typical Data Center Power Breakdown [3]

To the best of our knowledge, existing DCN energy saving solutions consider only either the hosts or the network, but not both. In this paper, we study the joint host-network optimization problem to improve the energy efficiency of DCNs. The basic idea is to simultaneously consider VM placement and network flow routing, so as to create more energy saving opportunities. The simplest way to combine host and network based optimization is just to naively first determine the VM placement and then the flow routing. Unfortunately, the existing VM placement algorithm [7] is not practical, since it does not consider the bandwidth capacity constraints of links, assumes fixed VM memory sizes, and has high time complexity of $O\,|V\,|^4$, where V is the set of VMs.. While our overarching research goal is to ultimately influence the next generation of router/switch hardware to make them more energy-aware, we would also like to introduce energy awareness in the operation of a large legacy base of equipment currently deployed. Thus we attempt to implement our algorithms with existing control knobs that are readily available in networking devices in operation today.

The rest of the paper is organized as follows. In Section II, we discuss related works. In Section III, we present a brief concept of depth-first best-fit search based algorithm. In Section IV, we present simulation results. In Section V, we conclude the paper.

Related Works

In this section, we briefly review existing energy saving solutions for DCNs and more broadly wide area networks. Those solutions can be divided into two broad categories: network-side optimization and host-side optimization.

A. Network-Side Optimization In the first category, ElasticTree [8] is a DCN power manager to find the set of switches and links that can accommodate the traffic and consume the minimum power. In addition, ElasticTree also addresses the robustness issue so that the optimized network has sufficient safety margins to prepare for traffic surges and network failures. GreenTE [9] manipulates the routing paths of wide area networks, so that the least number of routers shall be used to satisfy the performance constraints such as traffic demands and packet delays. Energy conservation can be achieved by then shutting down the idle routers and links without traffic. [10] proposes a energy saving scheme for the idle cables in bundled links.

B. Host-Side Optimization In the host-side optimization category, one approach is to optimize VM placement using live migrations [11], which will help consolidate VMs into fewer physical servers and traffic flows into fewer links. [12] proposes a traffic-aware VM placement scheme that localizes large traffic chunks and thus reduces load of high layer switches. The scheme achieves energy conservation by shutting down idle servers and switches after the placement. [13] studies the VM consolidation problem in the context of dynamic bandwidth demands. The problem is formulated as a stochastic bin packing problem and proved as NP-hard. The paper then proposes an approximation algorithm, which uses fewer servers while still satisfies all the performance constraints. The second host-side optimization approach is to improve the energy proportionality on the server itself. PowerNap [14] is an energy saving scheme for servers to quickly switch between two states: a high-performance active state to transmit traffic, and an idle state with low power to save energy.

We analyse the power consumption based on T.Ye et al [15]. The switch fabric architecture is constructed hierarchically. A network switch consists of four main parts: 1) the ingress packet process unit, 2) the egress packet process unit, 3) the arbiter (determines when and where a packet should be routed from the ingress ports to the egress ports) and 4) the switch fabrics is an interconnect network that connects the ingress ports to the egress ports. Power dissipation on switch can be categorized into two main parts as shown in Figure 1. 1) Internal buffer consumption ($E_{bit} = E_{access} + E_{ref}$) - The internal buffers, used to temporarily store the packets in buffer when contention between packets occurs. The less number of packets stored in buffers, the less power consumed. 2) Node switch power consumption ($E_{crossbar} = E_{NxN} + E_{sbit}$) - The internal node switches, located on the intermediate nodes between ingress and egress ports. They direct the packets from input ports to the next stage until reaching the destinations. The less number of packets sent, the less power used.

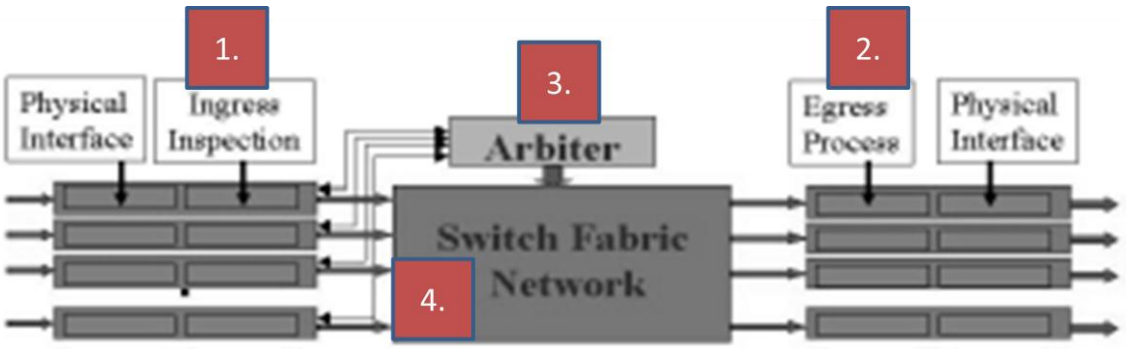

Figure 1. Switch Architecture [15]

Algorithm

It is easy to see that the joint host-network energy optimization problem is a variant of the multi-commodity problem [11], and can be formulated as a linear program. The optimization objective is to minimize the power consumption of all the servers, switches, and links in a DCN. Recent studies [12], [14], [16] indicate that power consumption of servers and switches in data centers can be roughly modeled as linear functions, which are suitable for linear programming. Even with non-linear power functions, various approximation techniques can help convert them to piece-wise linear ones as shown in Figure 2.

```
DFS(s, d, G) // s: source, d: destination, G: network
1   H = necessary-layer-to-connect(s, d, G);
2   path = {};
3   u = s;
4   next = 1; // flag indicating search direction, 1: upstream, -1: downstream
5   return SEARCH(u, path, next);

SEARCH(u, path, next) {
1   if (u = d) {path = path + u; return true;}
2   if ( layer-of(u) = H) next = −1; // reverse search direction after reaching connecting layer
3   if ( next = −1 && layer-of(u) = 1) return false;
4   neighbors = adjacent nodes of u in layer (layer-of(u) + next);
5   found = false;
6   while (neighbors ≠ ∅ && found = false) {
7      v = best-fit(neighbors); neighbors = neighbors − v;
8      found = SEARCH(v, path, next);
9   };
10  return found;
```

Figure 2. Depth-first Best-fit search based algorithm

In this paper, we first formulate the problem, and give the definitions of feasibility and fairness. Since integer linear programing is NP-complete, the above formulation is not suitable for practical deployment, but it can still be an ultimate bench mark to evaluate other approximation solutions.

Simulation results

First, besides performance and cost, another major issue that arises in data center design is power consumption. The switches that make up the higher tiers of the interconnect in data centers typically consume thousands of Watts, and in a large-scale data center the power requirements of the interconnect can be hundreds of kilowatts. Almost equally important is the issue of heat dissipation from the switches. Enterprise-grade switches generate considerable amounts of heat and thus require dedicated cooling systems.

In this section we analyze the power requirements and heat dissipation in our architecture and compare it with other typical approaches. We base our analysis on numbers reported in the switch data sheets, though we acknowledge that these reported values are measured in different ways by different vendors and hence may not always reflect system characteristics in deployment.

To compare the power requirement for each class of switch, we normalize the total power consumption and heat dissipation by the switch over the total aggregate bandwidth that a switch can support in Gbps.

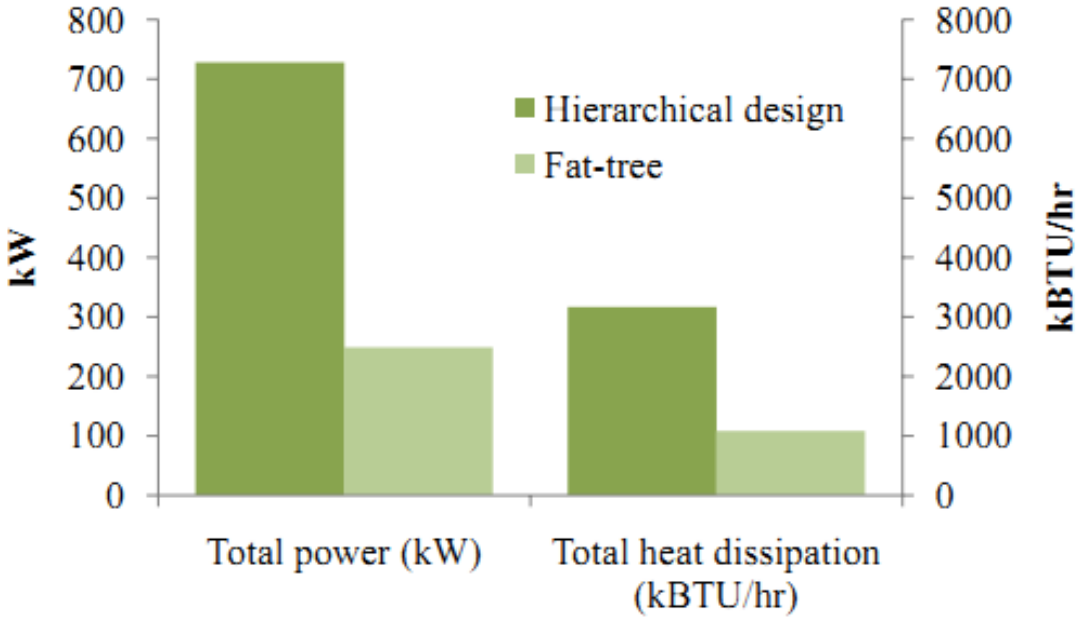

Figure 3. Comparison of total power consumption and heat dissipation.

Finally, we also calculated the estimated total power consumption and heat dissipation for an interconnect that can support roughly 27k hosts. Figure 3 shows that while our architecture employs more individual switches, the power consumption and heat dissipation is superior to those incurred by current data center designs, with 56.6% less power consumption and 56.5% less heat dissipation. Of course, the actual power consumption and heat dissipation must be measured in deployment; we leave such a study to our ongoing work.

Conclusions

Energy efficiency has become a high priority objective in most IT operational environments. Networks (including data center and enterprise networks) constitute an important part of the IT infrastructure and consume significant amounts of energy. Relatively little attention has been paid to improving the energy efficiency of networks thus far. Towards this end, in this paper, we make several contributions - (1) We propose a unified representation method to convert the virtual machine placement problem to a routing problem (2) We perform a parallelizing approach divides DCN into clusters and processes the clusters in parallel for fast completion (3) We quantify a fast topology oriented mutipath routing algorithm that can quickly find paths by using depth-first best-fit search based algorithm and (4) The simulation results show that our design is superior over existing host- or network-only optimization, and well approximates the ideal linear program. A more intelligent traffic routing using our algorithm scheme also yields significant network energy savings. We also incorporate service level awareness in our algorithms and show how network performance and redundancy can be traded off for energy savings. Our future work entails building a network power manager based on our findings and deploying it in a production network.

References

[1] [1] J. Li, L. Zhang, C. Lefurgy, R. Treumann, and W. E. Denzel, “Thrifty interconnection network for hpc systems,” ICS 2009: Proceedings of the 23rd International Conference on Supercomputing, pages 505–506, 2009.

[2] R. Raghavendra, P. Ranganathan, V. Talwar, Z. Wang, and X. Zhu, “No Power Struggles: A Unified Multi-level Power Management Architecture for the Data Center,” Proceedings Of ASPLOS, March 2008.

[3] P. Mahadevan, P. Sharma, S. Banerjee, and P. Ranganathan, “Energy aware network operations,” INFOCOM 2009: Proceedings of the 28th IEEE International Conference on Computer Communications Workshops, pages 25–30, 2009.

[4] A. Greenberg, J. Hamilton, D. Maltz, and P. Patel, “The cost of a cloud: Research problems in data center networks,” ACM SIGCOMM CCR: Editorial note, January 2009.

[5] R. N. Mysore, A. Pamboris, N. Farrington, N. Huang, P. Miri, S. Radhakrishnan, V. Subramanya, and A. Vahdat, “Portland: a scalable fault-tolerant layer 2 data center network fabric,” SIGCOMM Comput. Commun. Rev., 39(4):39–50, 2009.

[6] S. Nedevschi, J. Chandrashenkar, B. Nordman, S. Ratnasamy, and N. Taft, “Skilled in the Art of Being Idle: Reducing Energy Waste in Networked Systems,” Proceedings Of NSDI, April 2009.

[7] J. H. Ahn, N. Binkert, A. Davis, M. McLaren, and R. S. Schreiber, “HyperX: topology, routing, and packaging of efficient large-scale networks,” SC ’09: Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis, pages 1–11. ACM, 2009.

[8] J. Chabarek, J. Sommers, P. Barford, C. Estan, D. Tsiang, and S. Wright, “Power awareness in network design and routing,” Proceedings Of INFOCOM, April 2008.

[9] M. Al-Fares, A. Loukissas, and A. Vahdat, “A scalable, commodity data center network architecture,” SIGCOMM ’08: Proceedings of the ACM SIGCOMM 2008 Conference on Data Communication, 2008.

[10] G. Chen,W. He, J. Liu, S. Nath, L. Rigas, L. Xiao, and F. Zhao, “Energyaware server provisioning and load dispatching for connection-intensive internet services,” Proceedings Of NSDI, April 2008.

[11] Google Inc. Efficient computingâ step 2: efficient datacenters. http://www.google.com/corporate/green/datacenters/step2.html.

[12] D. Langen, A. Brinkmann and U. Ruckert, “High level estimation of the area and power consumption of on-chip interconnects,” IEEE Int’l ASIC/SOC Conference, 2000.

[13] A. Moustafa, M. Youssef, N. El-Derini and H. H. Aly, “Structure and Performance Evaluation of a Replicated Banyan Network Based ATM Switch” IEEE Sym. on Computers and Communications 1999.

[14] L. A. Barroso and U. Hölzle, “The case for energy-proportional computing,” Computer, 40(12):33–37, 2007. Mar. 2007.

[15] T. Ye, L. Benini and G. Micheli, "Analysis of Power Consumption on Switch Fabrics in Network Routers," ACM DAC 2002, New Orleans, LA, June 2002.

[16] P. Mahadevan, S. Banerjee and P. Sharma, "Energy Proportionality of an Enterprise Network," Green Networking, 2010.

Authors:

Tosmate Cheocherngngarn – received his B.Eng. degrees in Computer Engineering from Assumption University, Bangkok, Thailand in 2006. He is currently Ph.D. candidate at Electrical and Computer Engineering at Florida International University. His current research interests include Cross-layer approach for energy efficiency on data center network and High performance routers and switches.

Jean H. Andrian - received a diploma in mathematics from the University of Madagascar in 1975, a B.S. degree in Electrical Engineering and a M.S. degree in Engineering Physics from Ecole polytechnique de Montreal, Canada in 1979 and 1982 respectively, and a Ph.D. degree in Electrical Engineering from the University of Florida in 1985. Since 1985, he has been a faculty member of the Department of Electrical and Computer Engineering at Florida International University in Miami. His current research interests include digital signal processing in communications and theory and applications of wavelets to stochastic process.

Deng Pan – received his Master of Sciences and Bachelor of Science in Computer Science from Xi'an Jiaotong University, China, in 2002 and 1999, respectively; and also received his Ph.D. and M.S. in Computer Science from State University of New York at Stony Brook in 2007 and 2004, respectively. Since 2007, he has been a faculty member in FIU college of Engineering and Computing at Florida International University. His current research interests include High performance routers and switches, High speed networking, Quality of service, Network processors and Network security.

Khokiat Kengskool – received a Bachelor of Science degree in Industrial Engineering from Chulalongkorn University in Bangkok, Thailand in 1974, and Master Degree in Engineering Management from Missouri University of Science and Technology in 1976, and also Master and Ph.D. degrees in Industrial Engineering from the University of Missouri-Columbia in 1983 and 1986. Since 1986, he has been a faculty member in the Department of Industrial and Systems Engineering at Florida International University. His current research interests include Applied Artificial Intelligence, Decision-Making Support Systems and Productivity Enhancement.