An Energy-efficient Topological Framework for Data Center Networking

Narander Kumar and Swati saxena

Department of Computer Science, Babasaheb Bhimrao Ambedkar University (A Central University), Lucknow, India

Abstract: Energy efficiency in modern data centers is a necessity rather than a privilege today. As the internet traffic is growing exponentially every day, there is a mammoth amount of processing and communication workload which results in continuously active servers and switches. As a consequence, power consumption of a data center increases, thereby increasing its energy consumption. In order to make data center energy efficient, steps must be taken to limit the communication workload inside a data center and to switch off network elements when not in use. Various studies on green data centers suggest conserving energy of a server by migration, admission control, vm placements and so on. Few related works have discussed limiting the usage of data center switches. However, not much work can be found on data center topology and its effect on energy consumption. In this paper, we devise two models, an energy consumption model to outline the factors which are responsible for excessive energy consumption in a data center and an energy conservation model which tunes these responsible factors so that energy consumed is less. We apply these proposed models on two DCN topologies namely hierarchical fat-tree and recursive DCell to verify their utility. Our experimental work shows a significant improvement in energy conservation and as such is applicable to any DCN architecture.

Keywords: Data Center Networking; Energy efficiency; topological framework; three-tier architecture; DCell architecture

1. Introduction

Modern data centers are going through a phase of exponential growth in terms of its infrastructure and traffic. As a result, they experience a sharp escalation in energy consumption. Every element of a data center consumes power, like network switches, links, computing servers, storage etc. We are assuming power consumption to be a function of time, i.e., as long as a device is running/active, it is consuming power. Hence, the most direct step to reduce energy consumption is to deactivate a running device, be it switch or server or link. However, switching off a network element in a large data center is practically impossible because of huge and unpredictable traffic workload. A number of energy saving strategies have been offered in the past, most of which concentrate on reducing the energy consumption of computing servers as approximately 70% of a DCN energy consumption comes from them either in active and idle mode. However, proper consideration of network switches employed in a data center and their energy consumption pattern is crucial in maximizing the overall energy savings. In this paper, we have evolved an energy saving framework or model considering these two main elements of any data center, i.e., computing servers and network switches. Further this framework is applied to two popular DCN architectures, i.e., hierarchical layer-based switch-centric (fat tree) and recursive hybrid architectures (DCell) and the experimental results

The outline of the presented work is as follows- section 2 discusses the related work on energy efficiency in modern data centers. Section 3 introduces the energy consumption model in a large data center and explains the factors responsible for escalation in energy consumed. Section 4 discusses two example architectures of a DCN, three-tier and DCell architectures and their distinguishing features. Improvements in the underlying architectures are suggested in section 5 keeping in mind the goal of energy conservation. Section 6 provides relevant results obtained from implementing the presented models, followed by conclusion and scope of future enhancements.

2. Related Work

Cloud data centers enable customers to use computing services, platform and infrastructure with high efficiency and user-friendly billing system [1]. However, these data centers suffer from high computational cost due to increasing power and energy consumption [2]. This calls for the development of certain optimization techniques to handle and reduce the increase in energy consumption without adversely affecting the reliability and efficiency of data center resources like computing, storage, bandwidth, etc. [3]. As far as the energy consumption scenario is concerned, it is observed that IT and networking equipments consume nearly 50% of the total energy consumption of a data center [4]. Further, approximately half of such energy consumption is due to the data traffic inside a data center [5]. A large body of work, concerning energy efficiency in cloud data centers considers that datacenter infrastructures are underutilized [10] and over provisioned [3]. Among all the solutions offered, the Dynamic Power Management (DPM) method puts idle equipments into sleep mode [10], whereas Dynamic Voltage and Frequency Scaling (DVFS) [11] exploits the relation between power consumption *P*, supplied voltage V, and operating frequency *f* as reducing voltage and frequency reduces the power consumption. Also, power consumption of a server is linked with its CPU utilization and memory. As mentioned in [17] an idle server consumes about two-thirds of its peak power consumption. As far as the communication switches are concerned, their power consumption is partly fixed for chassis and line cards while energy consumed by the ports rise with the communication traffic.

Data replication models also help in optimization of data center energy [3], [6], [7] and [8].An extensive description of communication-aware models for cloud computing workloads is available in [23]. The most widely used data center topology is the three tier fat tree [16], which consists of three layers of network switches: core, aggregation and access. A comparatively newer topology, DCell, is surveyed in [24] which provide better results than fat tree architectures w. r. t. scalability and robustness. Most of the research work related to energy efficiency concentrate on load balancing and voltage-frequency trade-offs in switch-centric data center architecture. In this paper, a topological framework is proposed which suggests different energy saving solutions for two DCN architectures- three-tier and DCell by improving the storage features and communication pattern.

3. Energy Consumption Model

This section presents a framework to highlight the factors which effect the energy consumption in a DCN significantly. The presented model will serve as a guideline for improving the overall conservation of energy of a typical DCN. The two DCN elements worth considering in energy saving strategies are (i) computing servers which consume two-third of its peak power when sitting idle [27]and are responsible for approximately 70% of energy consumption in a DC and (ii) network switches which make up for approximately 20-30% of the total energy consumption in a DCN [29].

3.1 Energy consumption in a computing server- A typical modern data center consists of hundreds of thousands of computing servers communicating with each other through the network switches, usually arranged in a layered fashion, ex- Fat tree architecture. Each server consumes some fixed power even when its computing load is zero, referred here as idle power consumption or P_{idle}. Ideally, power consumed by a server depends on its load. At its maximum load, a server consumes maximum power which will be referred here as P_{max}. According to the related study [27], $P_{idle}$is equivalent to two-third of P_{max}. Keeping this equivalence in mind, the relation between power consumption of a server and its computing load can be stated as in equation 1.

$$P_{cs}(wl) = P_{idle} + (P_{\max} - P_{idle})\left(\frac{wl}{sc}\right) \quad [1]$$

where $P_{cs}(wl)$ is the power consumed by a computing server with a workload wl and sc is the total server capacity. Usually, (wl/sc)< 1. Substituting the relation between $P_{max}$and P_{idle}, one can conclude that

$$P_{cs}(wl) \cong \frac{P_{idle} * wl}{4} \quad [2]$$

Equation 2 further reinstates that power consumed by a server depends on its workload. Now, energy consumed by a computing server can be given as

$$E_{cs} = P_{cs}(wl) * T_{exec}(wl) \quad [3]$$

where E_{cs} is the energy consumption of a computing server and $T_{exec}(wl)$ is the total execution time of workload wl. Total execution time of a workload can be expressed as

$$T_{exec}(wl) = t_{proc} + 2 * t_{db} + t_{update} \quad [4]$$

Here, $t_{proc}$is the processing time of wl, $t_{db}$is the database access delay and t_{update} is the time required to update the data replicas. Processing time depends on the data size, database access delay depends on the location of the database in a DCN and updating time depends on the location of the concerned replica.

3.2 Energy consumption in a DCN switch- A typical DCN consists of thousands of switching elements responsible for inter-connection and routing among individual servers. As data communication increases inside a data center, network ports tend to be utilized to their maximum capacity. Energy consumed by a DCN switch is given in equation 5.

$$E_{sw} = P_{sw} * T \quad [5]$$

where E_{sw} is the energy consumed by a DCN switch, P_{sw} is the power consumed by a switch and T is the time a switch is active and is dependent on traffic flowing through it. Power consumed by a switch depends on the traffic passing through its ports [25, 26] and can be given as

$$P_{sw} = P_{const} + \sum_{q=1}^{m} U_q \quad [6]$$

where P_{const} is fixed power consumed by the switch's chassis and line cards, m are the number of ports in a switch and U_q is the throughput of a link associated with a port q.

4. DCN Architectures

A modern data center uses low-cost commodity servers and enterprise-class networking switches [28]. Since most of the data communication happens inside a data center, an energy-efficient DCN architecture must try to utilize limited number of switches for traffic propagation without affecting the service performance. As stated earlier, this paper considers two widely used DCN architectures, namely switch-centric (e.g., 3-tier architecture) and hybrid architecture (e.g., DCell) to apply the energy conservation model presented in the next section. We chose 3-tier architecture because it is the most commonly used DCN topology providing 1:1 oversubscription ratio. Due to scalability issues, many data centers are moving towards newer architectures, one of which is DCell having very good scaling features. Three-tier data center architecture is a hierarchical network of switches arranged in three layers, namely core, aggregation and access layer, as shown in figure 1.

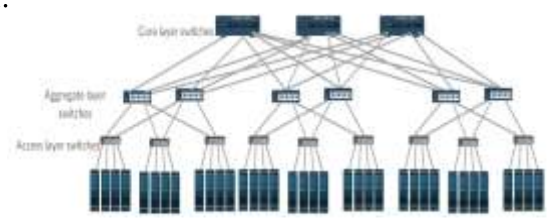

Fig. 1. Three-tier DCN architecture

As evident from the above figure, scalability may become a major issue in such data centers as growing data traffic will bottleneck the switches and will degrade the services' performance. Moreover, as large number of switches is engaged in communication, energy consumption will be more. Steps must be taken to activate as less number of switches as possible for any workload execution. DCell is a comparatively new server-centric hybrid DCN architecture where a single cell consisting of n servers and one switch ($DCell_0$) acts as a building block of the entire DCN network. $DCell_0$ forms level 0. At level 1, n+1 $DCell_0$ are required where each $DCell_0$ is connected to other DCells at the same level. Table I shows an example of the recursive nature of a typical DCell architecture.

Table 1. An example statistics of a DCell architecture

No. of Levels (k)	No. of DCells of lower levels (m)	No. of servers (n)
0	0	2
1	3	6
2	7	42
3	43	1806
:	:	:
k	$n_{k-1}+1$	$n_{k-1}*m$

In general, number of lower level DCells at level k are $\mathbf{n_{k-1}+1}$ and total number of servers at a level k are $\mathbf{n_{k-1}*m}$. This architecture enhances scalability, robustness and removes congestion bottleneck of three-tier architecture by recursively adding cells (or pods) to a DCN. Every server in DCell is equipped with network interface cards (NICs), thus enabling them with networking features. Different DCells at the same level are inter-connected via servers. Hence, DCell is hybrid architecture. Figure 2 shows DCell architecture at a recursive level 2.

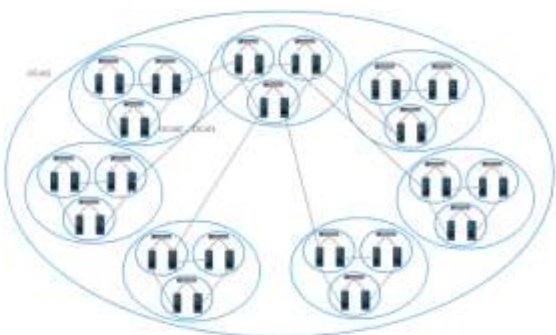

Fig. 2. Recursive 2-level DCell Architecture Example

A distinguishing feature of DCell is that it uses fewer switches as compared to three-tier architecture. However, this also means that servers in DCell are doing additional job of switching.

5. Energy Conservation Model

In this section, we investigate methods to reduce energy consumption in both the above mentioned DCN architectures by using the energy consumption model presented in section 1. As shown in equation 3, energy consumed by computing servers and switches depends on power consumed and time for which they are running. Energy conservation model described here emphasizes on reducing the workload time of servers and running time of switches.

5.1 Energy conservation in three-tier architecture- In order to reduce energy consumption in three-tier DCN architecture, we propose storage hierarchy. Traditionally a DCN accesses the central/main database located in a network cloud for every data access and update. This scenario requires data traffic to propagate all the way from a rack server to the main database and vice-versa every time, which increases traffic at all the three layers of switches. As a result, more power will be consumed at the intermediate switches, resulting in large energy consumption. To improve this scenario, we introduce rack databases at the access layer and data center database at the core layer network in addition to the original main database in the network cloud. Frequently accessed data can be stored in rack database to limit database transactions to the access layer. Likewise, lesser frequent data can be kept in DC database, to further restrict transactions within the data center. For rarely accessed data, contacting the main database will be required. This way a storage hierarchy, where databases are kept at three levels, will substantially reduce traffic at the aggregate and core level switches, resulting in less power consumption.

Fig. 3. Modified Three-tier Architecture

5.2 Energy conservation in DCell DCN architecture- The changes we propose in a recursive DCell architecture consist of reducing the number of active links connecting DCells at a single level by making redundant links sleep in cases of low traffic scenarios. As shown in figure 2, each $DCell_{l-1}$ is connected to every $DCell_{l-1}$ in $DCell_l$. This creates a mesh connectivity which increases redundancy and traffic interference at peak times. Moreover, during low traffic, all links are active which results in unnecessary consumption of power. In order to reduce the number of links, we suggest a modified DCell architecture as shown in figure 4 below. Every DCell is connected to its nearest neighbors only. Communication between two distant DCells at any level can take place via connecting DCells. Redundant links are put in sleeping mode which consumes much lesser power as compared to active links. By minimizing the total number of active links at any level in DCell architecture, we have reduced the routing load of each server in order to conserve its energy. More over in case of any active link failure, one of the redundant sleeping links can be activated so that there is no adverse effect on performance or reliability. Table 2 given below compares the number of active links at level 2 of the original and modified DCell architectures. Further, we also propose 'delayed forwarding' mechanism to be implemented in servers where update data will not be immediately sent to the database instead it will be delayed till non-peak traffic times, so as to reduce unnecessary traffic in the DCN during heavy traffic workload.

Table 2. Link minimization in DCell Architecture

Total no. of links at level 2 ($DCell_2$)	Original DCell Architecture	Modified DCell Architecture
	42	7

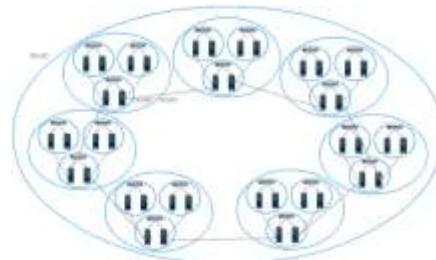

Fig. 4. Modified DCell Architecture Example showing active links

6. Simulation Results

Our simulation setup for three-tier architecture assumed uniform distribution of services and traffic among the servers and within the data center network. Parameters used for simulation are given in table 3.

Table 3. Three-tier Architecture parameters

Parameter	Value
Number of Core Switches	1
No. of aggregate switches	2
No. of access Switches	8
No. of servers	256
Idle power consumption of switches	150 W
Max power used by switches	2998 W

As summarized in table 3, our simulated data center consists of 256 computing servers distributed in 8 racks and inter connected by 2 aggregate and 1 core switches. Access links work at 1Gb/s while other links are at 10Gb/s. The idle and maximum power consumption of a computing server is considered to be 198 W and 300 W respectively. With the introduction of databases at three levels, energy consumption of computing servers was seen to be dipping as database access delay was greatly reduced.

Figure 5 shows the trend of energy consumption vs. access delay in 3-tier architecture. As seen, most of a server's requests for data were successfully fulfilled by the rack database, as a result lesser trips to the main or dc databases were needed. This reduced the routing load on aggregate and core level switches, hence energy consumed was reduced.

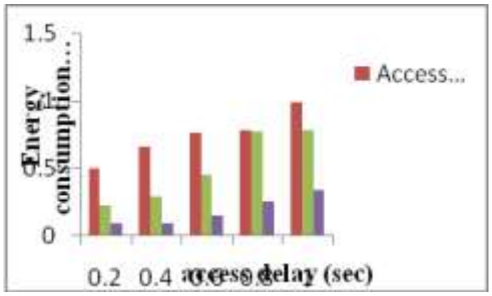

Fig. 5. Access delay vs energy consumption of switches in 3-tier DCN architecture

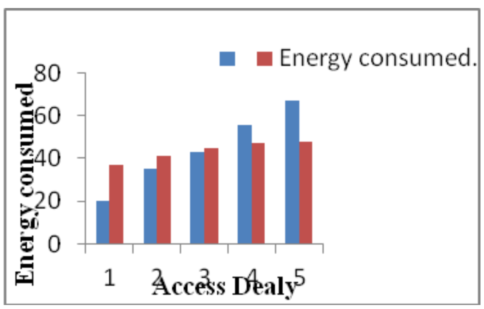

Fig. 6. Data access delay vs energy consumed in modified DCell

For DCell topology, we used 4 servers to make a $DCell_0$. Accordingly 5 $DCell_0$s were arranged to make $DCell_1$. Links connecting servers to switch are at 1Gb/s while inter-DCell links are at 10GB/s. Figure 6 shows the access delay and energy consumed in the proposed modified DCell architecture. With the reduction in data center active links and the introduction of 'delayed forwarding' mechanism, one can see that as data access delay increases, the energy consumption of the datacenter also increases. Modified DCell DCN architecture tries to reduce peak time traffic within the data center by implementing 'delayed forwarding' in routing servers. This delays the replica update operation to the non-peak traffic hours and thus, peak traffic time is dedicated to data access and processing operations. It helps in reducing unnecessary workload on servers which are already struggling with the dual responsibilities of computation and routing. Figure 7 shows a substantial reduction in energy consumption in the modified DCell architecture.

Figure 8 shows the energy consumption during packet loss in modified DCell architecture after introducing the 'delayed forwarding' mechanisms in servers for updates and backups.

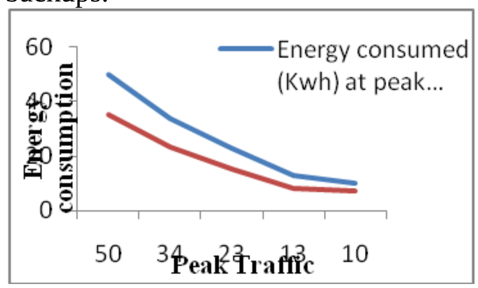

Fig. 7. Energy consumption during peak traffic in modified DCell

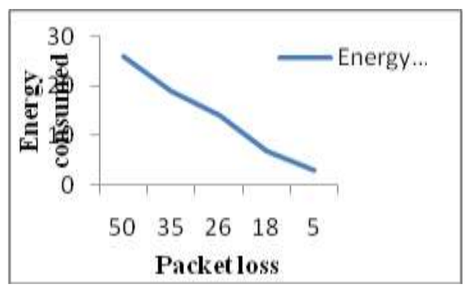

Fig. 8. Energy consumed during packet loss in modified DCell

7. Conclusion and Future Perspective

Exponentially increasing energy consumption by large data centers is one of the main concerns today. Measures to reduce power consumption require an understanding of the DCN architecture and its components. This paper examines two DCN architectures- three-tier fat tree and DCell for their energy consumption and factors affecting it. Three-tier architecture is very common in existing DC networks so fine tuning this architecture for energy reduction makes sense. On the other hand new DC networks are adopting newer architecture. DCell, being technically and structurally different from three-tier, it was chosen for implementation of energy conservation model. Modifications in the two architectures are proposed so that overall DCN's energy is conserved without adversely affecting performance. However, our proposed energy conservation framework works well for other DC architectures as well. Experimental results show a significant improvement in the overall energy consumption of a DC network. Future work involves further fine tuning the parameters responsible for energy consumption to establish a green DCN.

References

1. B. Hayes, "Cloud computing," Magazine Communications of the ACM, vol. 51, no. 7, pp. 9-11, 2008.
2. R. H. Katz, "Tech Titans Building Boom," IEEE Spectrum, vol. 46, no. 2, pp. 40-54, Feb. 2009.
3. Ruay-Shiung Chang, Hui-Ping Chang, and Yun-Ting Wang, "A dynamic weighted data replication strategy in data grids," IEEE/ACS International Conference on Computer Systems and Applications, pp. 414-421, April 2008.
4. Brown, R., et al.: Report to congress on server and data center energy efficiency: public law 109-431. Lawrence Berkeley National Laboratory, Berkeley, 2008.
5. Li Shang, Li-ShiuanPeh, and N. K. Jha, "Dynamic voltage scaling with links for power optimization of interconnection networks," Ninth International Symposium on High-Performance Computer Architecture (HPCA), pp. 91-102, February 2003.
6. Bin Lin, Shanshan Li, Xiangke Liao, Qingbo Wu, and Shazhou Yang, "eStor: Energy efficient and resilient data center storage," 2011 International Conference on Cloud and Service Computing (CSC), pp. 366-371, December 2011
7. Xiaowen Dong, T. El-Gorashi, and J. M. H. Elmirghani, "Green IP Over WDM Networks With Data Centers," Journal of Lightwave Technology, vol. 29, no. 12, pp. 1861-1880, June 2011.
8. Fan Ping, Xiaohu Li, C. McConnell, R. Vabbalareddy, Jeong-Hyon Hwang, "Towards Optimal Data Replication Across Data Centers,"International Conference on Distributed Computing SystemsWorkshops (ICDCSW), pp. 66-71, June 2011.
9. D. Kliazovich, P. Bouvry, and S. U. Khan, "GreenCloud: A Packet-levelSimulator of Energy-aware Cloud Computing Data Centers,"Journal of Supercomputing, vol. 62, no. 3, pp. 1263-1283, 2012.
10. Shengquan Wang, Jun Liu, Jian-Jia Chen, and Xue Liu,"PowerSleep: A Smart Power-Saving Scheme With Sleep forServers Under Response Time Constraint," IEEE Journal onEmerging and Selected Topics in Circuits and Systems, vol. 1, no. 3,pp. 289-298, September 2011.
11. T. Horvath, T. Abdelzaher, K. Skadron, and X. Liu, "Dynamicvoltage scaling in multitier web servers with end-to-end delaycontrol," IEEE Transactions on Computing, vol. 56, no. 4, pp. 444–458, 2007.
12. Li Shang, Li-ShiuanPeh, and N. K. Jha, "Dynamic voltage scalingwith links for power optimization of interconnection networks,"Ninth International Symposium on High-Performance ComputerArchitecture (HPCA), pp. 91-102, February 2003.
13. Xiaowen Dong, T. El-Gorashi, and J. M. H. Elmirghani, "Green IPOver WDM Networks With Data Centers," Journal of LightwaveTechnology, vol. 29, no. 12, pp. 1861-1880, June 2011.
14. Fan Ping, Xiaohu Li, C. McConnell, R. Vabbalareddy, Jeong-HyonHwang, "Towards Optimal Data Replication Across Data Centers,"International Conference on Distributed Computing SystemsWorkshops (ICDCSW), pp. 66-71, June 2011.
15. S. Asur, B. A. Huberman, G. Szabo, and C. Wang, "Trends in socialmedia: Persistence and decay," 5th International AAAI Conferenceon Weblogs and Social Media, 2011.
16. M. Al-Fares, A. Loukissas, and A. Vahdat, "A scalable, commoditydata center network architecture," ACM SIGCOMM, pp. 63-74,2008.
17. X. Fan, W.-D. Weber, and L. A. Barroso, "Power Provisioning for aWarehouse-sized Computer", ACM International Symposium onComputer Architecture, pp. 13-23, San Diego, CA, June 2007.

18. P. Mahadevan, P. Sharma, S. Banerjee, and P. Ranganathan, “Apower benchmarking framework for network devices,” 8th International IFIP-TC 6 Networking Conference, pp. 795 – 808,Aachen, Germany, May 2009.
19. P. Reviriego, V. Sivaraman, Z. Zhao, J. A. Maestro, A. Vishwanath,A. Sanchez-Macian, and C. Russell, “An energy consumption modelfor Energy Efficient Ethernet switches,” International Conference onHigh Performance Computing and Simulation (HPCS), pp. 98-104,July 2012.
20. D. Boru, D. Kliazovich, F. Granelli, P. Bouvry, A. Y. Zomaya,“Energy-Efficient Data Replication in Cloud ComputingData centers,” Springer Cluster Computing, 2014.
21. Intel Inc. (2010) Intel® Xeon® Processor 5000Sequence, available at: http://www.intel.com/p/en_US/products/server/processor/xeon5000
22. N. Farrington, E. Rubow, and A. Vahdat, “Data center switcharchitecture in the age of merchant silicon,” 17th IEEE symposiumon high performance interconnects (HOTI ’09), pp. 93–102, 2009.
23. D. Kliazovich, J. E. Pecero, A. Tchernykh, P. Bouvry, S. U. Khan,and A. Y. Zomaya, “CA-DAG:Communication-Aware Directed Acyclic Graphs for Modeling Cloud Computing Applications,”IEEE 6th International Conference on Cloud Computing (CLOUD), Santa Clara, CA, USA, 2013.
24. Bilal, Kashif, Samee Ullah Khan, Joanna Kolodziej, Limin Zhang, Khizar Hayat, Sajjad Ahmad Madani, Nasro Min-Allah, Lizhe Wang, and Dan Chen, "A Comparative Study Of Data Center Network Architectures," In *ECMS*, pp. 526-532. 2012.
25. D. Boru, D. Kliazovich, F. Granelli, P. Bouvry and A. Y. Zomaya, "Models for efficient data replication in cloud computing datacenters," 2015 IEEE International Conference on Communications (ICC), London, 2015, pp. 6056-6061.
26. PriyaMahadevan, Puneet Sharma, Sujata Banerjee, and ParthasarathyRanganathan. 2009. A Power Benchmarking Framework for Network Devices. In Proceedings of the 8th International IFIP-TC 6 Networking Conference (NETWORKING '09), Luigi Fratta, Henning Schulzrinne, Yutaka Takahashi, and Otto Spaniol (Eds.). Springer-Verlag, Berlin, Heidelberg, 795-808.
27. Xiaobo Fan, Wolf-Dietrich Weber, and Luiz Andre Barroso. 2007. Power provisioning for a warehouse-sized computer. In Proceedings of the 34th annual international symposium on Computer architecture (ISCA '07). ACM, New York, NY, USA, 13-23.
28. Bilal, Kashif, et al. "A Comparative Study Of Data Center Network Architectures." *ECMS*. 2012.
29. Li Shang, Li-ShiuanPeh, and Niraj K. Jha. 2003. Dynamic Voltage Scaling with Links for Power Optimization of Interconnection Networks. In Proceedings of the 9th International Symposium on High-Performance Computer Architecture (HPCA '03). IEEE Computer Society, Washington, DC, USA, 91-102, February 2003.