This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE Globecom 2010 proceedings.

GreenCloud: A Packet-level Simulator of Energy-aware Cloud Computing Data Centers

Dzmitry Kliazovich and Pascal Bouvry
FSTC CSC/SnT, University of Luxembourg
6 rue Coudenhove Kalergi, Luxembourg
dzmitry.kliazovich@uni.lu, pascal.bouvry@uni.lu

Yury Audzevich
University of Trento
Sommarive 14, 38100 Trento, Italy
audzevich@disi.unitn.it

Samee Ullah Khan
North Dakota State University
Fargo, ND 58108-6050
samee.khan@ndsu.edu

Abstract—**Cloud computing data centers are becoming increasingly popular for the provisioning of computing resources. The cost and operating expenses of data centers have skyrocketed with the increase in computing capacity. Several governmental, industrial, and academic surveys indicate that the energy utilized by computing and communication units within a data center contributes to a considerable slice of the data center operational costs.**

In this paper, we present a simulation environment for energy-aware cloud computing data centers. Along with the workload distribution, the simulator is designed to capture details of the energy consumed by data center components (servers, switches, and links) as well as packet-level communication patterns in realistic setups.

The simulation results obtained for two-tier, three-tier, and three-tier high-speed data center architectures demonstrate the effectiveness of the simulator in utilizing power management schema, such as voltage scaling, frequency scaling, and dynamic shutdown that are applied to the computing and networking components[1].

Keywords-energy efficiency, next geneartion networks, cloud computing simulations, data centers

I. INTRODUCTION

Over the last few years, cloud computing services have become increasingly popular due to the evolving data centers and parallel computing paradigms. The notion of a *cloud* is typically defined as a pool of computer resources organized to provide a computing function as a utility. The major IT companies, such as Microsoft, Google, Amazon, and IBM, pioneered the field of cloud computing and keep increasing their offerings in data distribution and computational hosting.

The operation of large geographically distributed data centers requires considerable amount of energy that accounts for a large slice of the total operational costs for cloud data centers. Gartner group estimates energy consumptions to account for up to 10% of the current data center operational expenses (OPEX), and this estimate may rise to 50% in the next few years [1]. However, computing based energy consumption is not the only power-related portion of the OPEX bill. High power consumption generates heat and requires an accompanying cooling system that costs in a range of \$2 to \$5 million per year for classical data centers.

Failure to keep data center temperatures within operational ranges drastically decreases hardware reliability and may potentially violate the Service Level Agreement (SLA) with the customers. A major portion (over 70%) of the heat is generated by the data center infrastructure; therefore, an optimized infrastructure installation may play a significant role in the OPEX reduction.

The first power saving solutions focused on making the data center hardware components power efficient. Techniques, such as Dynamic Voltage and Frequency Scaling (DVFS), and Dynamic Power Management (DPM) [2] were extensively studied and widely deployed. Because the aforementioned techniques rely on power-down and power-off methodologies, the efficiency of these techniques is at best limited. In fact, an idle server consumes about 2/3 of the peak load [3].

Because the workload of a data center fluctuates on the weekly (and in some case on hourly basis), it is a common practice to overprovision computing and communicational resources to accommodate the peak (or expected maximum) load. In fact, the average load accounts only for 30% of data center resources [4]. This allows putting the rest of the 70% of the resources into a sleep mode for most of the time. However, achieving the above requires central coordination and energy-aware workload scheduling techniques. Typical energy-aware scheduling solutions attempt to: (a) concentrate the workload in a minimum set of the computing resources and (b) maximizing the amount of resources that can be put into sleep mode. Moreover, performing power management dynamically during runtime considering wide range of system parameters may be up to 70% more efficient rather than static optimization [14].

Most of the current state-of-the-art research on energy efficiency has predominantly focused on the optimization of the processing elements. However, as recorded in earlier research, more than 30% of the total computing energy is consumed by the communication links, switching and aggregation elements. Similar to the case of processing components, energy consumption of the communication fabric can be reduced by scaling down the communication speeds and cutting operational frequency along with the input voltage for the transceivers and switching elements [5]. However, slowing the communicational fabric down should be performed carefully and based on the demands of user applications. Otherwise, such a procedure may result in a bottleneck, thereby limiting the overall system performance.

A number of studies demonstrate that often a simple optimization of the data center architecture and energy-aware

[1] The authors would like to acknowledge the support of Luxembourg FNR in the framework of GreenIT project (C09/IS/05).

scheduling of the workloads may lead to significant energy savings. Ref. [6] demonstrates energy savings of up to 75% that can be achieved by traffic management and workload consolidation techniques.

This article presents a simulation environment, termed GreenCloud, for advanced energy-aware studies of cloud computing data centers in realistic setups. GreenCloud is developed as an extension of a packet-level network simulator Ns2 [7]. Unlike the (only) existing cloud computing simulator CloudSim [8], GreenCloud extracts, aggregates, and makes information about the energy consumed by computing and communication elements of the data center available on an unprecedented fashion. In particular, a special focus is devoted to accurately capture communication patterns of currently deployed and future data center architectures.

The rest of the paper is organized as follows: Section II presents the main simulator components and related energy models; Section III focuses on the thorough evaluation of the developed simulation environment; Sefction IV concludes the paper providing the guidelines for building energy-efficient data centers and outlining directions for future work on the topic.

II. SIMULATION OF ENERGY-EFFICIENT DATA CENTER

A. Energy Efficiency

From the energy efficiency perspective, a cloud computing data center can be defined as *a pool of computing and communication resources organized in the way to transform the received power into computing or data transfer work to satisfy user demands*. Only a part of the energy consumed by the data center gets delivered to the computing servers directly. A major portion of the energy is utilized to maintain interconnection links and network equipment operations. The rest of the electricity is wasted in the power distribution system, dissipates as heat energy, and used up by air-conditioning systems. In light of the above discussion, in GreenCloud, we distinguish three energy consumption components: (a) computing energy, (b) communicational energy, and (c) the energy component related to the physical infrastructure of a data center.

B. Structure of the Simulator

GreenCloud is an extension to the network simulator Ns2 [7], which we developed for the study of cloud computing environments. The GreenCloud offers users a detailed fine-grained modeling of the energy consumed by the elements of the data center, such as servers, switches, and links. Moreover, GreenCloud offers a thorough investigation of workload distributions. Furthermore, a specific focus is devoted on the packet-level simulations of communications in the data center infrastructure, which provide the finest-grain control and is not present in any cloud computing simulation environment.

Fig. 1 presents the structure of the GreenCloud extension mapped onto the three-tier data center architecture.

C. Simulator components

Servers (S) are the staple of a data center that are responsible for task execution. In GreenCloud, the server components implement single core nodes that have a preset on a processing power limit, associated size of the memory/storage resources, and contains different task scheduling mechanisms ranging from the simple round-robin to the sophisticated DVFS- and DNS-enabled.

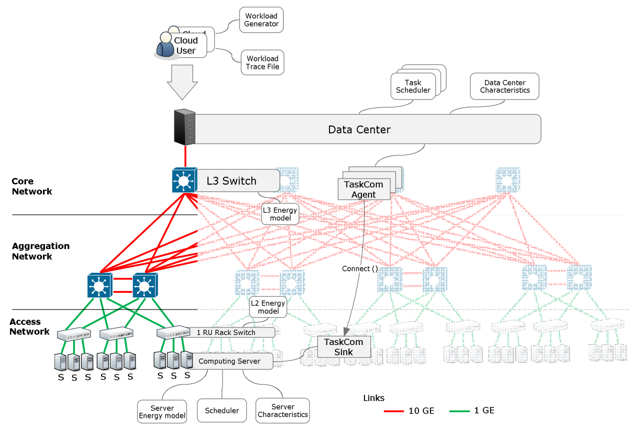

Figure 1. Architecture of the GreenCloud simulation environment.

The servers are arranged into racks with a Top-of-Rack (ToR) switch connecting it to the access part of the network.

The power model followed by server components is dependent on the server state and its CPU utilization. As reported in [3] an idle server consumes about 66% of its fully loaded configuration. This is due to the fact that servers must manage memory modules, disks, I/O resources, and other peripherals in an acceptable state. Then, the power consumption increases with the level of CPU load linearly. As a result, the aforementioned model allows implementation of power saving in a centralized scheduler that can provision the consolidation of workloads in a minimum possible amount of the computing servers.

Another option for power management is Dynamic Voltage/Frequency Scaling (DVFS) [5], which introduces a tradeoff between computing performance and the energy consumed by the server. The DVFS is based on the fact that switching power in a chip decreases proportionally to $V^{2*}f$. Moreover, voltage reduction requires frequency downshift. This implies a cubic relationship from f in the CPU power consumption. Note that server components, such as bus, memory, and disks do not depend on the CPU frequency. Therefore, the power consumption of an average server can be expressed as follows [11]:

$$P = P_{fixed} + P_f * f^3 \quad (1)$$

Fig. 2 presents the server power consumption model implemented in GreenCloud. The scheduling depends on the server load level and operating frequency, and aims at capturing the effects of both of the DVFS and DPM techniques.

Switches and Links form the interconnection fabric that delivers workload to any of the computing servers for execution in a timely manner.

The interconnection of switches and servers requires different cabling solutions depending on the supported bandwidth, physical and quality characteristics of the link. The quality of signal transmission in a given cable determines a tradeoff between transmission rate and the link distance, which

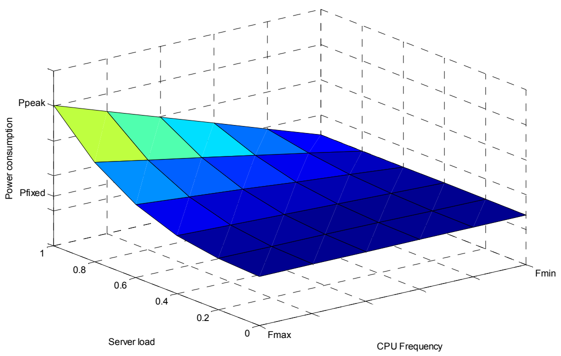

Figure 2. Computing server power consumption.

are the factors defining the cost and energy consumption of the transceivers.

The twisted pair is the most commonly used medium for Ethernet networks that allows organizing Gigabit Ethernet (GE) transmissions for up to 100 meters with the transceiver power consumed of around 0.4W or 10 GE links for up to 30 meters with the transceiver power of 6W.

The twisted pair cabling is a low cost solution. However, for the organization of 10 GE links it is common to use optical multimode fibers. The multimode fibers allow transmissions for up to 300 meters with the transceiver power of 1W. On the4 other hand the fact that multimode fibers cost almost 50 times the twisted pair cost motivates the trend to limit the usage of 10 GE links to the core and aggregation networks as spending for the networking infrastructure may top 10-20% of the overall data center budget [12].

The number of switches installed depends on the implemented data center architecture. However, as the computing servers are usually arranged into racks the most common switch in a data center is Top-of-Rack (ToR) switch. The ToR switch is typically placed at the top unit of the rack unit (1RU) to reduce the amount of cables and the heat produced. The ToR switches can support either gigabit (GE) or 10 gigabit (10GE) speeds. However, taking into account that 10 GE switches are more expensive and current capacity limitation of aggregation and core networks gigabit rates are more common for racks.

Similar to the computing servers early power optimization proposals for interconnection network were based on DVS links [5]. The DVS introduced a control element at each port of the switch that depending on the traffic pattern and current levels of link utilization could downgrade the transmission rate. Due to the comparability requirements only few standard link transmission rates are allowed, such as for GE links 10 Mb/s, 100 Mb/s, and 1Gb/s are the only options.

On the other hand, the power efficiency of DVS links is limited as only a portion (3-15%) of the consumed power scales linearly with the link rate. As demonstrated by the experiments in [13] the energy consumed by a switch and all its transceivers can be defined as:

$$P_{switch} = P_{chasis} + n_{linecards} \cdot P_{linecard} + \sum_{i=0}^{R} n_{ports.r} \cdot P_r \,, \qquad (2)$$

where $P_{chassis}$ is related to the power consumed by the switch hardware, $P_{linecard}$ is the power consumed by any active network line card, P_r corresponds to the power consumed by a port (transceiver) running at the rate r. In Eq. (2), only the last component appears to be dependent on the link rate while other components, such as $P_{chassis}$ and $P_{linecard}$ remain fixed for all the duration of switch operation. Therefore, $P_{chassis}$ and $P_{linecard}$ can be avoided by turning the switch hardware off or putting it into sleep mode. This fact motivated a combination of DVS scheme with DNS (dynamic network shutdown) approach.

The proposed GreenCloud simulator implements energy model of switches and links according to Eq. (2) with the values of power consumption for different elements taken in accordance as suggested in [6]. The implemented powers saving schemes are: (a) DVS only, (b) DNS only, and (c) DVS with DNS.

Workloads are the objects designed for universal modeling of various cloud user services, such as social networking, instant messaging, and content delivery. The execution of each workload object requires a successful completion of its two main components: (a) computational and (b) communicational.

The computational component defines the amount computing resourced required to be provided by the computing server in MIPS or FLOPS and the duration in time for which these computing resources should be allocated.

The communicational component of the workload defines the amount and the size of data transfers that must be performed before, during, and after the workload execution. It is composed of three parts: (a) the size of the workload, (b) the size of internal, and (c) the size of external to the data center communications.

The size of the workload defines the number of bytes that being divided into IP packets are required be transmitted from the core switches to the computing servers before a workload execution can be initiated.

The size of external communications defines the amount of data required to be transmitted outside the data center network at the moment of task completion. However, the internal communications account for the workloads scheduled at different servers that have interdependencies. The internal communications specify the amount of data to be communicated with a randomly chosen server inside the data center at the moment of task completion. In fact, internal communication in the data center can account for as much as 70% of total data transmitted [6].

An efficient and effective methodology to optimize energy consumption of interdependent workloads is to analyze the workload communication requirements at the moment of scheduling and perform a coupled placement of these interdependent workloads – a co-scheduling approach. The co-scheduling approach will reduce the number of links/switches involved into communication patterns.

The workload arrival rate/pattern to the data center can be configured to follow a predefined distribution, such as Exponential or Pareto, or can be re-generated from traces log files. Furthermore, the trace-driven workload generation is designed to simulate more realistic workload arrival process capturing also intraday fluctuations [4], which may influence simulated results greatly.

III. PERFORMANCE EVALUATION

In this section we present case study simulations of an energy-aware data center for two-tier (2T), three-tier (3T), and three-tier high-speed (3Ths) architectures.

For comparison reasons we fixed the number of computing nodes to 1536 for all three topologies while the number and interconnection of network switches varied. Table I summarizes the main simulation setup parameters.

In contrast with other architectures 2T data center does not include aggregation switches. The core switches are connected to the access network directly using 1 GE links (referred as C_2-C_3) and interconnected between them using 10 GE links (referred as C_1-C_2).

The 3Ths architecture mainly improves the 3T architecture with providing more bandwidth in the core and aggregation parts of the network. The bandwidth of the C_1-C_2 and C_2-C_3 links in 3Ths architecture is ten times of that in 3T and corresponds to 100 GE and 10 GE respectively. The availability of 100 GE links allows keeping the number of core switches as well as the number of paths in ECMP routing limited to 2 serving the same amount switches in the access. The propagation delay of all the links is set to 10 ns.

The task generation events and the size of the tasks are exponentially distributed with an average task size fixed at 4500 bytes which corresponds to 3 Ethernet packets.

The tasks arrived to the data center are scheduled for execution using energy-aware "green" scheduler. This "green" scheduler tends to group the workload on a minimum possible amount of computing servers. The servers left idle are put into a sleep mode (DNS) while on the under-loaded servers the supply voltage is reduced (DVFS).

Fig. 3 presents a workload distribution among severs. The whole load of the data center (around 30% of its total capacity) is mapped onto approximately one third of the servers maintaining load at a peak rate (left part of the chart). This way, the remaining two thirds of the servers can be shut down using DNS technique. A tiny portion of the approximately 50 out of 1536 servers which load represents a falling slope of the chart are under-utilized on average, and DVFS technique can be applied on them.

Table II presents the power consumption of data center components. The server peak energy consumption of 301 W is composed of 130 W [2] (43%) allocated for a peak CPU consumption and 171 W (56%) consumed by other devices like memory, disks, peripheral slots, mother board, fan, and power supply unit [10]. As the only component which scales with the load is the CPU power, the minimum consumption of an idle server is bounded and corresponds to 198 W (66%) where also a portion of CPU power consumption of 27 W required to keep the operating system running is included.

The switches' consumption is almost constant for different transmission rates as most of the power (85-97%) is consumed by their chassis and line cards and only a small portion (3-15%) is consumed by their port transceivers. Switch power consumption values are derived from [6] with a twisted pair cable connection considered for the rack switch (C_3) and optical multimode fiber for the core (C_1) and aggregation (C_2) switches.

TABLE I. SIMULATION SETUP PARAMETERS

Parameter	Data center architectures		
	Two-tier	**Three-Tier**	**Three-tier high-speed**
Topologies			
Core nodes (C_1)	16	8	2
Aggregation nodes (C_2)	-	16	4
Access switches (C_3)	512	512	512
Servers (S)	1536	1536	1536
Link (C_1-C_2)	10 GE	10 GE	100 GE
Link (C_2-C_3)	1 GE	1 GE	10 GE
Link (C_3-S)	1 GE	1 GE	1 GE
Link propagation delay	10 ns		
Data center			
Data center average load	30%		
Task generation time	Exponentially distributed		
Task size	Exponentially distributed		
Average task size	4500 bytes (3 Ethernet packets)		
Simulation time	60 inutes		

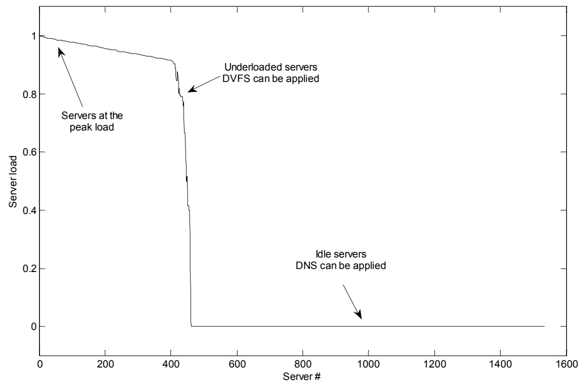

Figure 3. Server workload distribution with a "green" scheduler.

Table III presents simulation results obtained for three evaluated data center topologies. On average, the data center consumption is around 503kW·h during an hour of the runtime. On the yearly basis it corresponds to 4409MW·h or $441k with an average price of 10c per kW·h.

The processing servers share around 70% of total data center energy consumption while the communicational links and switches account for the rest 30% of the total amount. Furthermore, the consumption of switches breaks with 17% allocated for core switches, 34% for aggregation switches, and 50% for the access switches. This way, also taking into account the requirements for network performance, load balancing, and communication robustness the obvious choice is to keep core and aggregation switches constantly running at the full speed. Fig. 4 reports an average distribution of energy consumption in a data center.

Table IV compares the impact on energy consumption of DVFS, DNS, and DVFS with DNS schemes applied on both computing several and networking equipment. The DVFS scheme alone reduces power consumption to 84% from the nominal level. Most of the reduction comes from downshifting CPU voltage as CPU components accounts for 43% of the total energy consumed by the server. On the other hand DVFS shows itself ineffective for the switches as only a tiny portion switch's energy (3%) is sensitive to the transmission rate variation.

[2] Chosen based on the specification of Intel Xeon 4-core processor with 8MB of cache running at 3.33 GHz.

TABLE II. POWER CONSUMPTION OF DATA CENTER COMPONENTS

Parameter	Power consumption (W)		
Servers			
Server peak	301		
Server CPU peak	130		
Server other (memory, peripherial, mother board, fan, PSU losses)	171		
Server idle	198		
Switches			
	Access network (C_3)	Core (C_1) and Aggregation (C_2)	
Chassis	146	1.5K (10G)	15K (100G)
Linecard	-	1K (10G)	12K (100G)
Port tranceiver	0.42	0.3K (10G)	1.6K (100G)

TABLE III. DISTRIBUTION OF DATA CENTER POWER CONSUMPTION

Parameter	Power consumption (kW·h)		
	Two-tier	*Three-Tier*	*Three-tier high-speed*
Data center	477.8	503.4	508.6
Servers	351	351	351
Switches	126.8	152.4	157.6
Core (C_1)	51.2	25.6	56.8
Aggregation (C_2)	-	51.2	25.2
Access (C_3)	75.6	75.6	75.6

TABLE IV. COMPARISON OF ENERGY-EFFICIENT SCHEMES

Parameter	Power consumption (kW·h)			
	No energy-saving	DVFS	DNS	DVFS+DSS
Data center	503.4	486.1 (96%)	186.7 (37%)	179.4 (35%)
Servers	351	340.5 (97%)	138.4 (39%)	132.4 (37%)
Switches	152.4	145.6 (95%)	48.3 (32%)	47 (31%)

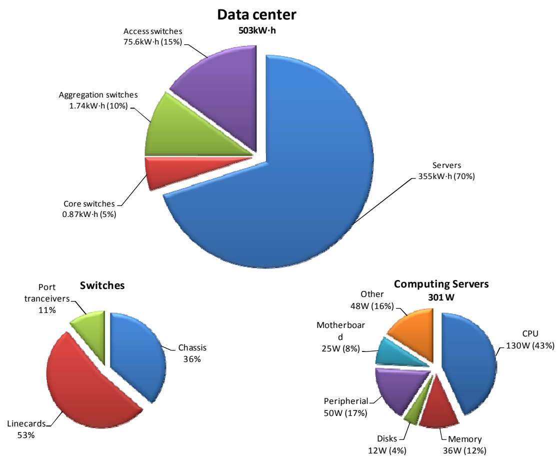

Figure 4. Distribution of energy consumption in a data center.

The most effective results are obtained by DNS scheme. It is equally effective for both servers and switches as the most of their energy consumed shows no dependency on the operating frequency. However, in order to utilize DNS scheme effectively its design should be coupled with the data center scheduler positioned to unload the maximum number of the servers.

It should be noted that due to the limited size (4500 bytes) of the tasks generated by the cloud users the impact of traffic patterns on the interconnection network became minimized. This fact in part led to the similarity of energy consumed by different data center architectures. The effects of variable task sizes and dense traffic loads in the interconnection network will be explored in the future works on the topic.

IV. CONCLUSIONS

In this paper we presented a simulation environment for energy-aware cloud computing data centers. Greencloud is designed to capture details of the energy consumed by data center components as well as packet-level communication patterns between them.

The simulation results obtained for two-tier, three-tier, and three-tier high-speed data center architectures demonstrate applicability and impact from the application of different *power management schemes like voltage scaling or dynamic* shutdown applied on the computing as well as on the networking components.

The future work will focus on the simulator extension adding storage area network techniques and further refinement of energy models used in the simulated components. On the scheduling part the analysis will be focused on optimal task allocation with cooperative scheduling techniques [9].

REFERENCES

[1] Gartner Group, available at: http://www.gartner.com/

[2] T. Horvath, T. Abdelzaher, K. Skadron, and Xue Liu, "Dynamic Voltage Scaling in Multitier Web Servers with End-to-End Delay Control," IEEE Transactions on Computers, vol. 56, no. 4, pp. 444 – 458, 2007.

[3] G. Chen, W. He, J. Liu, S. Nath, L. Rigas, L. Xiao, and F. Zhao, "Energy-aware server provisioning and load dispatching for connection-intensive internet services," the 5th USENIX Symposium on Networked Systems Design and Implementation, Berkeley, CA, USA, 2008.

[4] J. Liu, F. Zhao, X. Liu, and W. He, "Challenges Towards Elastic Power Management in Internet Data Centers", Workshop on Cyber-Physical Systems (WCPS), Montreal, Quebec, Canada, June 2009.

[5] Li Shang, Li-Shiuan Peh, and Niraj K. Jha, "Dynamic Voltage Scaling with Links for Power Optimization of Interconnection Networks," Symposium on High-Performance Computer Architecture, 2003.

[6] P. Mahadevan, P. Sharma, S. Banerjee, and P. Ranganathan, "Energy Aware Network Operations," IEEE INFOCOM, pp. 1 – 6, 2009.

[7] The Network Simulator Ns2, avialable at: http://www.isi.edu/nsnam/ns/

[8] R. Buyya, R. Ranjan, and R. N. Calheiros, "Modeling and Simulation of Scalable Cloud Computing Environments and the CloudSim Toolkit: Challenges and Opportunities," 7th High Performance Computing and Simulation Conference, Leipzig, Germany, June, 2009.

[9] S. U. Khan and I. Ahmad, "A Cooperative Game Theoretical Technique for Joint Optimization of Energy Consumption and Response Time in Computational Grids," IEEE Transactions on Parallel and Distributed Systems, vol. 21, no. 4, pp. 537-553, 2009.

[10] X. Fan, W.-D. Weber, and L. A. Barroso, "Power provisioning for a warehouse-sized computer," ISCA, New York, NY, USA, 2007.

[11] Y. Chen, A. Das, W. Qin, A. Sivasubramaniam, Q. Wang, and N. Gautam, "Managing server energy and operational costs in hosting centers," ACM SIGMETRICS, New York, USA, pp. 303-314, 2005.

[12] A. Greenberg, P. Lahiri, D. A. Maltz, P. Patel, and S. Sengupta, "Towards a next generation data center architecture: scalability and commoditization," ACM Workshop on Programmable Routers For Extensible Services of Tomorrow, Seattle, WA, USA, August, 2008.

[13] P. Mahadevan, P. Sharma, S. Banerjee, and P. Ranganathan, "A Power Benchmarking Framework for Network Devices," IFIP-TC 6 Networking Conference, Aachen, Germany, May, 2009.

[14] B. Khargharia, S. Hariri, F. Szidarovszky, M. Houri, H. El-Rewini, S. U. Khan, I. Ahmad, and M. S. Yousif, "Autonomic Power and Performance Management for Large-Scale Data Centers," IEEE International Parallel and Distributed Processing Symposium (IPDPS), Long Beach, CA, USA, March 2007.