Semantic Analysis of energy consumption for estimating SLA violation with virtual machine provisioning schema using DVFS

Arunadevi[a], Dr. R.S.D Wahidabanu[c]
[a]Assistant Professor of Sambhram Academy of Management Studies, Bangalore, India
Mobile: 9538202730
[b]Professor and Head of Department of ECE of Government College of Engineering, Salem, India
Mobile: 9443008886

Abstract: Energy consumption is one of the major expense in IT industry which has attracted the attentions of various data centers which deploys huge quantity of power supply. So attempt to reduce energy utilization is very prime concern in cloud computing as it supports various complex services obeying SLA. In this research work, we present an in-depth analysis of energy consumption at a core model of cloud computing in order to visualize the risk factors which is responsible for draining energy in data centers. We create an energy analysis which will consist of various cloudlet-IDs with estimation of SLA violation along with virtual machine migration with complete status capturing. Various energy host will be designed as real time virtual machine along with provisioning of various processing elements (PE) of datacenters using dynamic voltage frequency scaling methodology. The simulation will be designed in java environment to show the performance of our analysis approach to reduce energy consumption.

Keywords: Energy consumption, Cloud Computing, Data center.

I. INTRODUCTION

When the computing services are supplied by various real time datacenters over the internet can be commonly termed as cloud computing. Cloud computing assures virtually limitless computational resources to its users, while letting them pay only for the resources they actually use at any given time. Various cloud computing services such as Amazon EC2 [1] and Google App Engine [2] are designed to take benefit of the already existing infrastructure of their respective business which delivers computing services to users as a utility in a pay-as-you-go manner [3]. The different services facilitated by the cloud providers are Platform as a Service (PaaS), Infrastructure as a Service (IaaS), and finally Software as a Service (SaaS). In reality, Service providers make high-quality use of IaaS and PaaS for developing their services without consideration of physical hardware, while users also can access on-demand and pay-per-use services anywhere in Cloud computing.

But one of major issue in datacenters found is to manage optimum energy usage in the systems. It was found that there is a consumption of 10 to 100 times high energy per square foot compared to distinctive office building for an average scale datacenters [4]. They can even consume as much electricity as a whole city [5]. The majority of energy consumption in datacenters approaches from computation processing, disk storage, network, and various types of cooling systems. In this research proposal, we will deploy virtual machine provisioning to analyze various energy processing functionalities along with their parameters. This can be one of the noteworthy technique in cloud computing for evaluation of energy models in Cloud computing

The concept of cloud computing guarantees Service Level Agreements (SLAs) between customers and Cloud suppliers with Pay-as-you-go mechanism which specify that the negotiated agreements as deadline. Therefore, every datacenters will attempt to minimize energy consumption without violating these SLAs. As various real-time applications will require deadline constraints, this research proposal will concentrate on power-aware real-time Cloud services, such as distributed image processing, realtime distributed databases, financial analysis and so on. The main contribution of this proposed system will be to give analysis and in-depth study of real-time Cloud service with virtualization, and (ii) to probe various energy-aware virtual machine optimizing schemes based on Dynamic Voltage Frequency Scaling schemes.

The remainder of this paper is organized as follows. Section II presents problem statement followed by related works in Section III. The proposed approach is discussed in Section IV. Section V will highlight the methodology deployed in this research work followed by Simulation work in Section VI and finally Section VII will conclude the research proposal.

II. PROBLEM DESCRIPTION

Cloud computing has been accounted for diversified emerging issues which is yet to be resolved. Unfortunately, the current potentials and capability of cloud computing has yet not been enhanced to cater the crucial scope of the usage. There are technical and nontechnical gaps. One of the most prominent issues is that although vision of cloud computing is to deploy at any location irrespective of time and devices, particularly in case of distributed system and existing clouds is reported to deliver very poor performance in applications. Any significant issue is the migration of data from client's infrastructure to clouds actually cost very high and it is also a time consuming process. Various legal restrictions are yet to be created for the acceptance of this new technology in socioeconomic market. In many data centers today, the cost of power is only second to the actual cost of the equipment investments. Today the cloud data center consumes 1-2 percent of world energy. If left unchecked, they will rapidly consume an ever-increasing amount of power consumption

III. RELATED WORK

Dara Kusic et. al. [6] implement and validate a dynamic resource provisioning framework for virtualized server environments wherein the provisioning problem is posed as one of sequential optimization under uncertainty and solved using a look ahead control scheme. A new suite for placement and energy consolidation of the virtual machines in data centers has been discussed by Michael Cardosa [7]. The author has validated the results obtained from experiments conducted in artificial and real data centers testbeds and concluded that the protocol designed constantly provides the optimal throughput on a large spectrum of inputs. A unique design and accomplishment of an architecture for managing resources over the hosting center operating system has been proposed by Jeffrey S. Chase [8]. The prime importance of the work was the power for running the resource management problems for huge clusters of servers intended to provide automation in adapting to prescribed load for server resources. This has resulted in enhancement of the power efficiency for clusters of server by resizing the set of active servers dynamically and reacting to the supply of energy interference or thermal events by corrupting the service as per Service Level Agreements (SLAs).

A power-aware protocol which automatically acclimatizes its potential difference and frequency configuration is already proposed by Chunghsing Hsu [9]. The main intention is to facilitate considerable reduction in energy and power cutback with optimal effect on its respective performance. Kyong Hoon Kim [10] proposed power-aware scheduling algorithms for bag-of-tasks applications with deadline constraints on DVS-enabled cluster systems. The proposed scheduling algorithms select appropriate supply voltages of processing elements to minimize energy consumption. Linwei Niu [11] investigated the problem of applying scheduling techniques to reduce both the dynamic and leakage energy consumption. Monfort [12] proposed a hierarchical real-time virtual resource model which is ideal for the open system environment with a clean separation of concerns between task group scheduling and partition scheduling across multiple levels of resource decomposition. Anshul Gandhi [13] experimentally find that the power-to-frequency relationship within a server for a given workload can be either linear or cubic. Interestingly, we see that the relationship is linear for DFS and DVFS when the workload is CPU bound, but cubic when it is more memory bound. By contrast, the relationship for DVFS+DFS is always cubic in their experiments. Leping Wang [14] presents a threshold-based method for efficient power management of heterogeneous soft real-time clusters. An illustrative analysis for recognizing the environment for isolative applications was studied by Akshat Verma [15]. The author has also shown that energy utilization by various HPC applications can be dependent on application or non-linear and can posses huge diversified range. The inter-association of power utilization and workload performance has been analyzed by Shekhar Srikantaiah [16] which elicited the power performance trade-offs for consolidation and concludes that superior operating point do persists.

IV. PROPOSED SYSTEM

The proposed system will use cloud resource broker to initiate a role in finding various power list of various processing elements as well as virtual machine. The diversified factors in consideration will be energy consumption for real-time applications, VM with millions instructions per second (MIPS), and life-time of energy host. Thus, it is assumed that real-time virtual machines are defined from multiple real-time applications of the service by using compositional realtime technique. Therefore, virtual machine provisioner in our Clouds will map for virtual machines for the service and not for individual applications. The parameters under consideration are DVFS enabled group of cloudlets, task allocations to VMs with various performance models. The proposed system is assumed to consist of numerous nodes which are assumed to be operated on different voltage with different frequencies. The operating point is defined as $OP_i=f$ (Voltage$_{OP}$, Frequency$_{op}$) at ith operating point. Therefore it can be said that maximum processor operating voltage is product of final operating point and processor operating voltage and minimum voltage as product of initial operating point and voltage. This analysis model focuses on CPU power saving in terms of virtual machine provisioning in the cloud computing. The proposal will only consider the dynamic power dissipation because it is more dominating cause in the total power consumption. And, datacenters can augment their profit by dropping dynamic energy consumption.

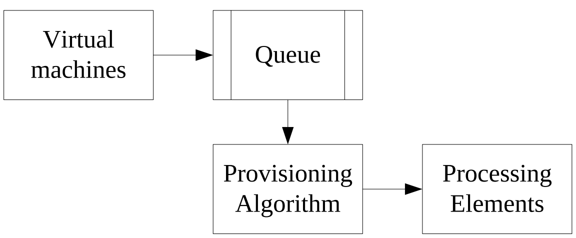

Fig.1. Proposed steps for provisioning

Fig.1 represents the input of service request from the virtual machines to the group of Processing Elements using DVFS enabled Provisioning Algorithm. The proposed algorithm executed on the group of PEs with specified gap-time. At the instant of provisioning, the VMs request is forwarded at the provisioner, which will be allocated for the consecutive pass that executes on the specified operating point. The system initiates with all the group of processing elements executing at the minimum processor speed and operating voltage. During the instant when the VM will accomplish its execution during the previous provisioning pass, the provisioner initially scales down energy report of the processing element. Exactly after this, the provisioner forwards the VMs request to the queue thereby reducing the energy consumption. The introduction of “gap-time” in the system model allows the provisioner to be in passive mode during the provisioning instant. The system then arranges all the request from VMs in descending sequence of necessary operating frequency. The system will be designed to provision VM with abundant energy required with maximum priority. In case the virtual machine is not provisioned in the previous pass, then they will be configured to the highest priority. The next stage of the system will be to minimize quantity of PE unutilized supplied voltage. Thereby the proposed system assures better energy efficiency without influencing the performance of the VMs.

V. METHODOLOGY

The proposed system has focused on group of numerous nodes which are known as processing elements or PE. The research methodology considers the analyzation of similar set of groups with all entities (PEs) to posses uniform processor speed and provide exactly similar processing performance (MIPS). The mathematical definition of this group “G” can be represented as:

$$G = \sum_{n=1}^{N} PE_n$$

Where PE_n is the nth processing element inside G with N-total number of processing element where PE_n can be represented as f (Operating Point$_{PE}$, Voltage$_{PE}$, Frequency$_{PE}$). Similarly, group of virtual machine “VM” could be mathematically represented as:

$$VM = \sum_{i=1}^{I} vm_i$$

Where vm_i represents ith virtual machine to be provisioned among I-total number of virtual machine, where vm_i can be represented as f(Speed$_{Req}$, Time$_{EXEC}$, Time$_{START}$). The proposed system is focused on dynamic energy consumption where these PE and VM will be considered. The Dynamic Voltage Frequency Scaling scheme reduces the dynamic energy consumption by decreasing the supplying voltage and frequency, which results in slowdown of the execution time. The energy consumption can be reduced by combining DVS (Dynamic Voltage Scaling) and proportional sharing scheduling. The datacenter will evaluate the schedulability for the processing elements of rate of MIPS. Since the provisioning policy is to provide lower price to users, the algorithm finds the minimum-price processor. For the same price, less energy is preferable because it produces higher profit. Finally, a virtual machine is mapped on allocated PE if virtual machine is schedulable on the datacenter. The protocol for effective provisioning of VM along with optimization of energy utilization can be done by reducing processor operating supply voltage which will be feasible by scaling down the processor frequency. It can also be done by provisioning VM to PE only with lower value of voltage without any tendency to scale for upper grade of voltage. The algorithm description are as below:

Algorithm: Provisioning VM using DVFS
Input: Virtual machines, Processing elements
Output: Effective provisioning of VM to PE.
Steps:
1. *Create function for VM*
2. *Design a queue*
3. *Initialize a Set of nodes.*
4. *Define variables for PE.*
5. **FOR** *n=1* to *N*
6. **DO**
7. *Estimate minimum processor operating frequency*
    PE_n . $Frequency_{PE}$
8. *Estimate maximum processor operating frequency.*
9. *Use PEn for maximum Voltage level*
10. *Estimate available processor speed*
11. **END**
12. *Initialize Operating time instant as 0*

//Scaling VM voltage down in provisioning pass
13. **FOR** *i=1 to I* **DO**
14. *Check if VM have completed their execution*
15. *For lower voltage*
16. *For PE, whose VM is free.*
17. *Update Available processor speed*
18. *Update Maximum processor speed*
19. *For higher voltage*
20. *Scale down PEn to operating point with maximum reduced voltage*

//Provisioning VM in gap-time
21. *Arrange all the VMs for their necessary processor speed*
22. VM_i . $Speed_{Req}$
23. **FOR** *p=1 to p<I*
24. **DO**
25. **FOR** *q=1 to j<I*
26. **DO**
27. *Search all PE in every q pass*
28. **IF** (PE_n . $Speed_{Req}$)>VM_i . $Speed_{Req}$ **THEN**
29. *Provision* VM_i *to* PE_n.
30. Iterate for all nodes
31. **END**
32. **END**
33. **IF** *VMi is not provisioned* **THEN**
34. *Search PEn with high operating voltage*
35. *Find lowest feasible voltage level>current voltage level.*
36. *Initiate provisioning of* VM_i *to* PE_n.
37. *Update network.*
38. *Substitute time instant by adding gap-time.*
39. *Iterate for all nodes.*
40. *Check complete Provisioning*

VI. CASE STUDY

The case study considered in this is designed on GridSim where large scale of independent task is considered to be processed. The SLA consideration of the proposed case study is maintained by configuring the deadline of the allocated task. The huge scale of global network is designed with geographical grouping of various cloudlets or gridlets in distributed order. A service broker is designed in order manage the task of controlling the routing of user request between data center.

The case study will consider energy consumption provisioning of multiple task allocation in DVFS enabled cloud as shown in Fig 2.

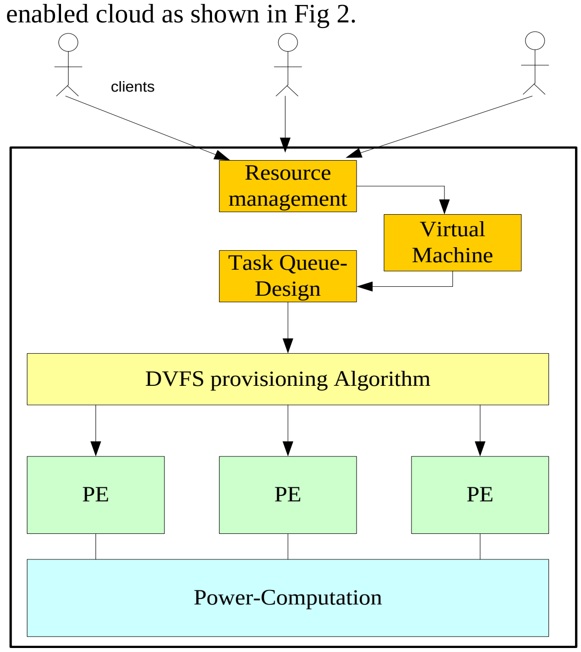

Fig. 2. System architecture of the case study.

The parameters used in proposed system is mapped to the case study using the similar components of VM, queue, DVFS algorithm, and PE. The case study shown is simulated with various nodes ranging from 10-50 where the VM is simulated in range of 100-500. The experiment is performed in Intel Pentium 1.84 GHz processor.

As shown in Fig 2, the proposed provisioning algorithm has potential to minimize energy utilization in DVFS enabled group of components. The energy utilization is directly proportional to quantity of VMs considered assuming constant quantity of processing elements. The energy utilization is inversely proportional to quantity of PE considered assuming constant quantity of VMs as seen in Fig 3.

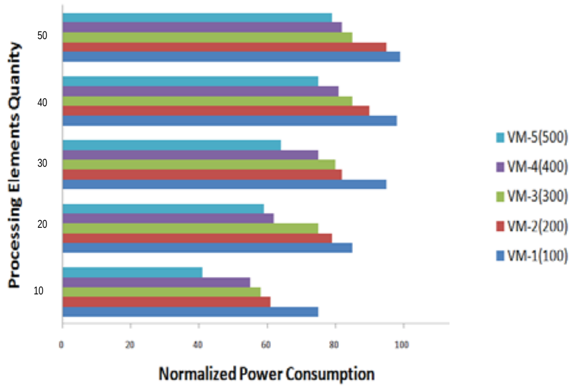

Fig 3. Proposed case study with DVFS algorithm

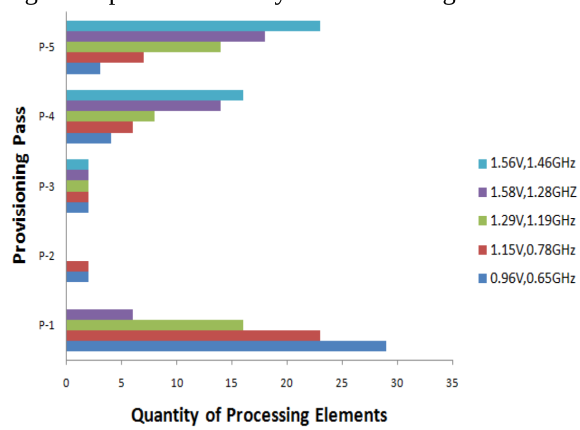

Fig 4. PE number Vs Provisioning pass

Fig 4. represents the allotment of operating frequencies considering provisioning pass and quantity of PE. The simulation is conducted considering lower to higher values of operating frequency and voltage. It can be seen that quantity of PE operating at 1.4 GHz maximizes and quantity of PE reduces when operating at 0.6 GHz.

VII. SIMULATION RESULT

The proposed system discussed is evaluated on java platform in 32 bit machine with windows OS of 1.84 GHz processor. The prototype design of DVFS protocol is layered into task-allocation, group creation, and processing elements. The virtual machines are developed for accepting number of task considering simulation time and processing speed of the system. The groups of processing elements are represented by the configuration of nodes and provisioning protocol designed while the proposed DVFS protocol is checked for its performance in the processing element. The system designs a datacenter with user specified machines considering 16 DVFS enabled processors of which the properties are as follows:

Table 2: Properties of datacenter

	# of PEs	MIPS of PE	DVFS level	α (10^{-3})
Station 1	5	1,784	[0, 1.0]	2.92
Station 2	5	2,500	[0, 1.0]	4.08
Station 3	5	2,987	[0, 1.0]	5.37
Station 4	5	3,501	[0, 1.0]	6.21

The simulation design model conducted will be almost equivalent to the Amazon Elastic Compute Cloud [2] where $0.10 is assumed for unit price per hour. The entire result analysis has considered the data shown in Table 2 where 500 VMs are involved in the analysis. The cumulative service quantity of VMs is arbitrarily chosen from 2,400 GIs to 3,600 GIs. The task completion time limit is chosen from 10-30 minutes additional than the implementation time depending on 1000 times of measure of computer processor speed. Discrete probability distribution is used for estimating inter-arrival time between two VMs. For the purpose of analysis of DVFS enabled design, the current work considers 3 different energy aware policies e.g. Minimized DVFS, Sigma advanced DVFS, and adjusted DVFS. Minimized DVFS renders the processor speed to be reduced to the score by which the VM should meet the task deadline. Sigma advanced DVFS will be used to consider the feasibility of maximizing the inward request of the VMs. Sigma value of constant depending on the load towards processing the task in system. Adjusted DVFS policy controls the mean arrival rate, mean service rate, and mean task completion deadline and thereby adjust the speed of the processor. Consideration of static is also made which operated on maximum processor speed where it doesn't implement DVFS. The performance analysis results are as below:

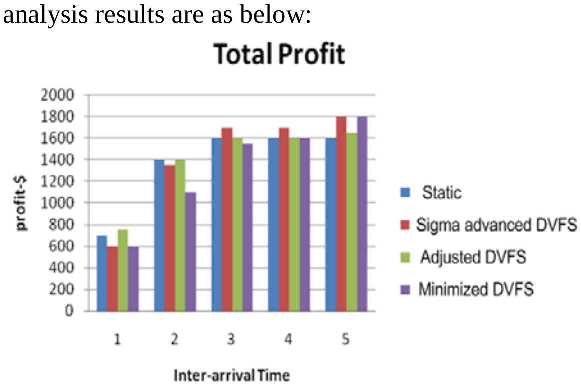

Fig 5. Inter-arrival Time vs profit

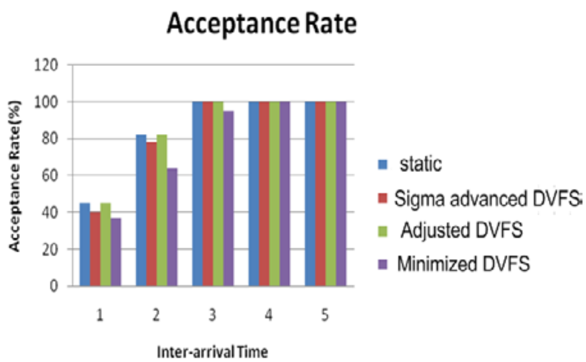

Fig 6. Inter-arrival time Vs Acceptance rate

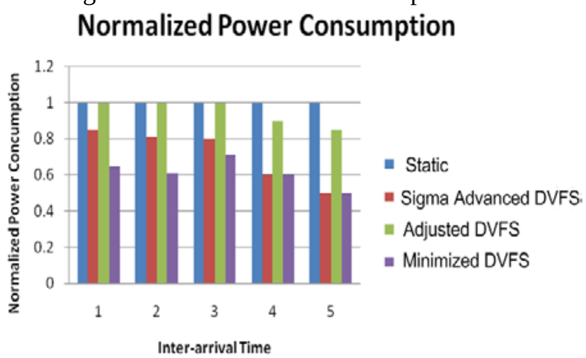

Fig 7. Inter-arrival time Vs. Normalized Power Consumption

Figure 5 represents the total profits of each considered policy in accordance with inter-arrival time. In sigma-advanced-DVFS, the value of sigma is considered 15%. It can be seen from the simulation results that policy considering without DVFS yields to greater profit as it considers all VMs. It also shows that adjusted DVFS policy has less profit when compared with policy without DVFS, but at the same time, it is also evident that other DVFS techniques represents higher profit in minimum inter-arrival rate for optimal energy consumption. Figure 6 and 7 represents acceptance rate and the normalized power consumption compare to Static, respectively for the VMs. The acceptance rate of adjusted-DVFS is very much equivalent to Static technique but minimizes much power in case of small rate of arrival. Sigma-advanced-DVS has maximum acceptance rate with uniform power utilization when compared to minimized-DVFS approach. Normally, sigma-advanced-DVFS exhibits the superlative performance with respect to profit per energy consumption as the quantity of scaling high is managed involuntarily in accordance with the load of task processing in system. In case of adjusted-DVFS, it can be seen that the performance is restricted by the simplified queueing model.

VIII. CONCLUSION

The research proposal probes various aspects of modeling the power consumption in data centers. The proposed system attempts to investigate a complex service analysis for cloud computing which is structured as virtual machine for various resource brokers. The research also finds various power-aware policies of virtual machines for Cloud services. Simulation environment is created in java platform which shows that datacenters can predominantly reduce energy consumption and increase their profit using proposed DVFS schemes which shows maximized profit with minimized power consumption irrespective of system load.

Reference

[1] “Amazon Elastic Compute Cloud,” http://aws.amazon.com/ec2.

[2] “Google App Engine,” http://code.google.com/appengine/.

[3] M. Armbrust, et al. Above the Clouds: A Berkeley view of cloud computing. Tech. Report No. UCB/EECS-2009-28, University of California at Berkeley, USA, February 2009

[4] P. Scheihing. Creating energy efficient data centers. In Data Center Facilities and Engineering Conference. Washington, DC, USA, May 2007.

[5] J. Markoff and S. Lohr. Intel’s huge bet turns iffy. New York Times Technology Section, September 2002.

[6] D. Kusic, et al. Power and performance management of virtualized computing environments via lookahead control. In Proc. of 5th IEEE Intl. Conf. on Autonomic Computing (ICAC 2008), pages 3–12. Chicago, USA, June 2008.

[7] M. Cardosa, M. R. Korupolu, and A. Singh. Shares and utilities based power consolidation in virtualized server environments. In Proc. of IFIP/IEEE Intl. Symp. On Integrated Network Management. USA, June 2009.

[8] J. S. Chase, et al. Managing energy and server resources in hosting centers. In Proc. of 8th ACM Symp. on Operating Systems Principles. Banff, Canada, October 2001.

[9] C. Hsu and W. Feng. A power-aware run-time system for high-performance computing. In Proc. of ACM/IEEE Conf. on Supercomputing. Seattle, USA, November 2005.

[10] K. H. Kim, R. Buyya, and J. Kim. Power aware scheduling of bag-of-tasks applications with deadline constraints on DVS-enabled clusters. In Proc. of 7th IEEE Intl. Symp. On Cluster Computing and the Grid (CCGrid’07), pages 541–548. Rio de Janeiro, Brazil, May 2007.

[11] L. Niu and G. Quan. Reducing both dynamic and leakage energy consumption for hard real-time systems. In Proc. Of CASES'04. Washington, DC, USA, Sept. 2004
[12] Valérie Monfort, Maha Khemaja, Nouha Ammari, Fayssal Fehli, Using SaaS and Cloud computing For "On Demand" E Learning Services, 2010 10th IEEE International Conference on Advanced Learning Technologies
[13] A. Gandhi, M. Harchol-Balter, R. Das, and C. Lefurgy. Optimal power allocation in server farms. In Proc. of Intl. Joint Conf. on Measurement and Modeling of Computer Systems, pages 157–168. Seattle, USA, June 2009.
[14] L. Wang and Y. Lu. Efficient power management of heterogeneous soft real-time clusters. In Proc. of IEEE Real-Time Systems Sym. Barcelona, Spain, Dec. 2008.
[15] A. Verma, P. Ahuja, and A. Neogi. Power-aware dynamic placement of HPC applications. In Proc. of ICS'08, pages 175–184. Agean Sea, Greece, June 2008.
[16] S. Srikantaiah, A. Kansal, and F. Zhao. Energy aware consolidation for cloud computing. In Workshop on Power Aware Computing and Systems (HotPower '08). San Diego, USA, December 2008.