Deepali Koul *et al*, International Journal of Computer Science and Mobile Applications,
Vol.1 Issue. 5, November- 2013, pg. 57-64 **ISSN: 2321-8363**

Analysis of Various Energy Efficient Data Centers Using Cloud Systems

DEEPALI KOUL
KAUTILYA COLLEGE OF ENGINEERING & TECHNOLOGY
JAIPUR

deep.koul19@gmail.com

CHAITALI PATEL
KAUTILYA COLLEGE OF ENGINEERING & TECHNOLOGY
JAIPUR

chaitali.er@gmail.com

***Abstract*--The Computing systems basically focus on the performance issues driven by factors as consumer demand, business need or scientific reasons. As Energy efficiency is increasingly important for future information and communication technologies (ICT), the increased usage of ICT, together with increasing energy costs and the need to reduce green house gas emissions call for energy-efficient technologies that decrease the overall energy consumption of computation, storage and communications. At the same time we also need to maintain QoS(quality of service) and SLA(service level agreement). For the same we need to identify challenges in the area of energy and power affected. After studying some of the current best practice and relevant literature in this area, this paper identifies some of the remaining key research challenges that arise when such energy-saving techniques are extended for use in cloud computing environments. We discuss causes and problems of high power/energy consumption, and present taxonomy of energy efficient design of computing systems covering the virtualization, and data centre levels. We survey various key works in the area and map them onto our taxonomy to guide future design and development efforts.**

1. INTRODUCTION

Cloud Computing is an umbrella term for how IT is delivered as a Service. Within the Cloud Computing domain, there are various terms for specific pieces of the cloud. Some of these terms include the different variations of “XaaS” – meaning “something” as a service; the X is a variable that can be changed to represent several things such as Storage as a Service, Security as a Service and others[1].Energy-efficient resource management is considered in parallel and distributed computing domain in recent years as Green computing. Cloud computing, which is popular with “pay-as-you-go” utility model is economy driven. Apart from the overwhelming operational cost, building a Data Centre leads to excessive establishment expenses as data centres are usually built to serve infrequent peak loads resulting in low average utilization of the resources.Moreover , there are other crucial problems that arise from high power consumption. Insufficient or malfunctioning Cooling system can lead to overheating of the resources reducing system reliability and devices lifetime. In addition high power consumption by the infrastructure leads to substantial carbon dioxide(Co_2) emissions leading to the Green house effect[2].

Energy consumption is not only determined by hardware efficiency, but it is also dependent on the resource management system deployed on the infrastructure and the efficiency of applications running in the system. Higher power consumption results not only in boosted electricity bills but also in additional requirements to a cooling system and power delivery infrastructure, that is, uninterruptible power supplies (UPS), power distribution units (PDU), and so on. With the growth of computer components density, the cooling problem becomes crucial, as more heat has to be dissipated for a square meter. Apart from the overwhelming operating costs and the total cost of acquisition (TCA), another rising concern is the environmental impact in terms of carbon dioxide (CO_2) emissions caused by high energy consumption. Therefore, the reduction of power and energy consumption has become a first-order objective in the design of modern computing systems.

Where is power consumed?

The bifurcation of power consumption in one Server is shown in Fig.1[3]. It shows that CPU quadcore consumes the highest power.

IJCSMA
Deepali Koul *et al*, International Journal of Computer Science and Mobile Applications,
Vol.1 Issue. 5, November- 2013, pg. 57-64 **ISSN: 2321-8363**

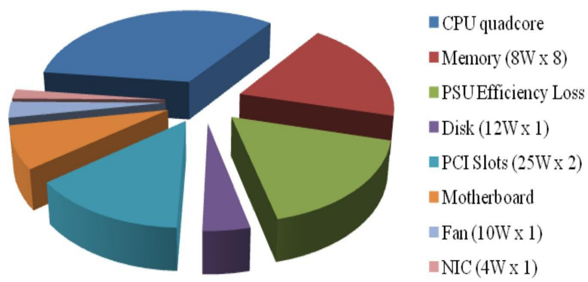

Figure.1 Power Consumption in one Server

Following measures indicates the average power consumption in one data centre.

Computer Rm. AC	34%
Server/Storage	50%
Power Conversion	7%
Network	7%
Lighting	2%

2. MODELS FOR POWER CONSUMPTION

A. STATIC POWER CONSUMPTION

The static power consumption, or leakage power, is caused by leakage currents that are present in any active circuit, independently of clock rates and usage scenarios. This static power is mainly determined by the type of transistors and process technology. The reduction of the static power requires improvements of the low-level system design.

.

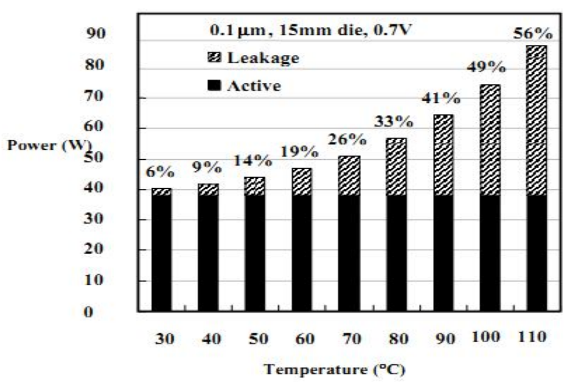

Figure.2 Power vs. Temperature comparison

For example, the leakage power consumption. On a typical ASIC in a modern nanometre process, the leakage power consumption cannot keep independent. It is much related to the temperature as shown in Fig.2 [3].

B. DYNAMIC POWER CONSUMPTION

It is created by circuit activity (transistor switches, changes of values in register, etc).and depends mainly on system's status (usage scenario, clock rates and IO activity. It is defined as equation mentioned below [3].

Deepali Koul *et al*, International Journal of Computer Science and Mobile Applications,
Vol.1 Issue. 5, November- 2013, pg. 57-64 **ISSN: 2321-8363**

$$P_{dynamic} = a \cdot C \cdot V^2 \cdot f$$

Where, a: switching activity
C: equivalent capacitance
V: supply voltage
f: clock frequency

C. MODELLING POWER CONSUMPTION

There is a strong relationship between the CPU utilization and server total power consumption. Power consumption by a server grows linearly with the growth of the CPU utilization from the value of the power consumption in the idle state up to the power consumed when the server is fully utilized. It is given by [3]

$$P(u) = P_{idle} + (P_{busy} - P_{idle}) \cdot (2u - u^r).$$

u: CPU's utilization
r: calibration parameter calculated practically

3. TAXONOMY OF POWER MANAGEMENT IN COMPUTING SYSTEM

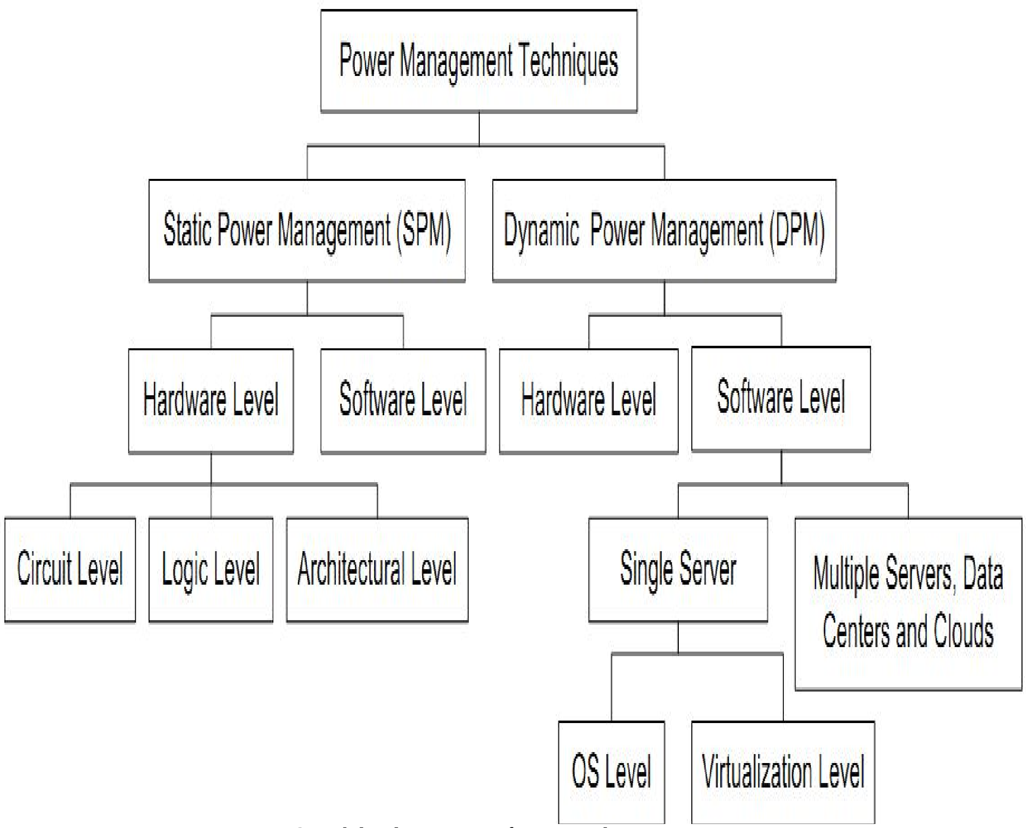

Figure.3 High level taxonomy of Power and Energy Management

Static power management mainly focuses on optimizing the circuit, logic and architecture of the system at the design time. Dynamic power Management includes methods and strategies for run-time adaptation of a system's behaviour according to current resource requirements or any other Dynamic characteristics of the system state. Another technology that can improve the utilization of resources, and thus reduce the power consumption, is virtualization of computer resources. The virtualization technology allows one to create several VMs on a physical server and, therefore, reduce the amount of hardware in use and improve the utilization of resources.Fig.3 [3], illustrates various level where power management can be possible. The concept originated with the IBM mainframe OSs of the 1960s, but was commercialized for x86-compatible computers only in the 1990s [3].

4. DYNAMIC VOLTAGE AND FREQUENCY SCALING

Dynamic voltage and frequency scaling is a commonly-used technique to save power on a wide range of computing systems, from embedded, laptop and desktop systems to high-performance server-class systems. DVFS is able to reduce the power consumption of a CMOS integrated circuit, such as a modern computer processor, by reducing the frequency at which it operates, as shown by

$$P = CfV_2 + P_{static}$$

Where C is the capacitance of the transistor gates (which depends on feature size), f is the operating frequency and V is the supply voltage. The voltage required for stable operation is determined by the frequency at which the circuit is clocked, and can be reduced if the frequency is also reduced. This can yield a significant reduction in power consumption because of the V_2 relationship shown above [4].

5. VIRTUALIZATION LEVEL

The virtualization level enables the abstraction of an OS and applications running on it from the hardware. Physical resources can be split into a number of logical slices called VMs. Each VM can accommodate an individual OS creating for the user a view of a dedicated physical resource and ensuring the performance and failure isolation between VMs sharing a single physical machine. The virtualization layer lies between the hardware and OS and, therefore, a virtual machine monitor (VMM) takes the control over resource multiplexing and has to be involved in the system's power management. There are two ways of how a VMM can participate in the power management [3]:
1. A VMM can act as a power-aware OS without distinction between VMs: monitor the overall system's performance and appropriately apply DVFS or any DCD techniques to the system components.
2. Another way is to leverage OS's specific power management policies and application-level knowledge, and map power management calls from different VMs on actual changes in the hardware's power state or enforce system-wide power limits in a coordinated manner.
Most popular virtualization technology solutions: are the Xen hypervisor, VMware solutions, and Kernel-based virtual machine (KVM).

6. DATA CENTRE LEVEL

Data centre level taxonomy is based on the idea of consolidating the workload into the minimum of physical resources. Switching off idle resources leads to the reduced energy consumption, as well as the increased utilization of resources; therefore, lowering the TCO(Total Cost of Ownership) and speeding up Returns On Investments (ROI).However, to meet the SLA requirements, the consolidation has to be done intelligently in order to minimize both the energy consumption and performance degradation.
There are different approaches to addressing the problem of effectively managing the energy-performance trade-off in virtualized and non-virtualized data centres. The characteristics used to classify the approaches are presented in Fig. 4 [3].

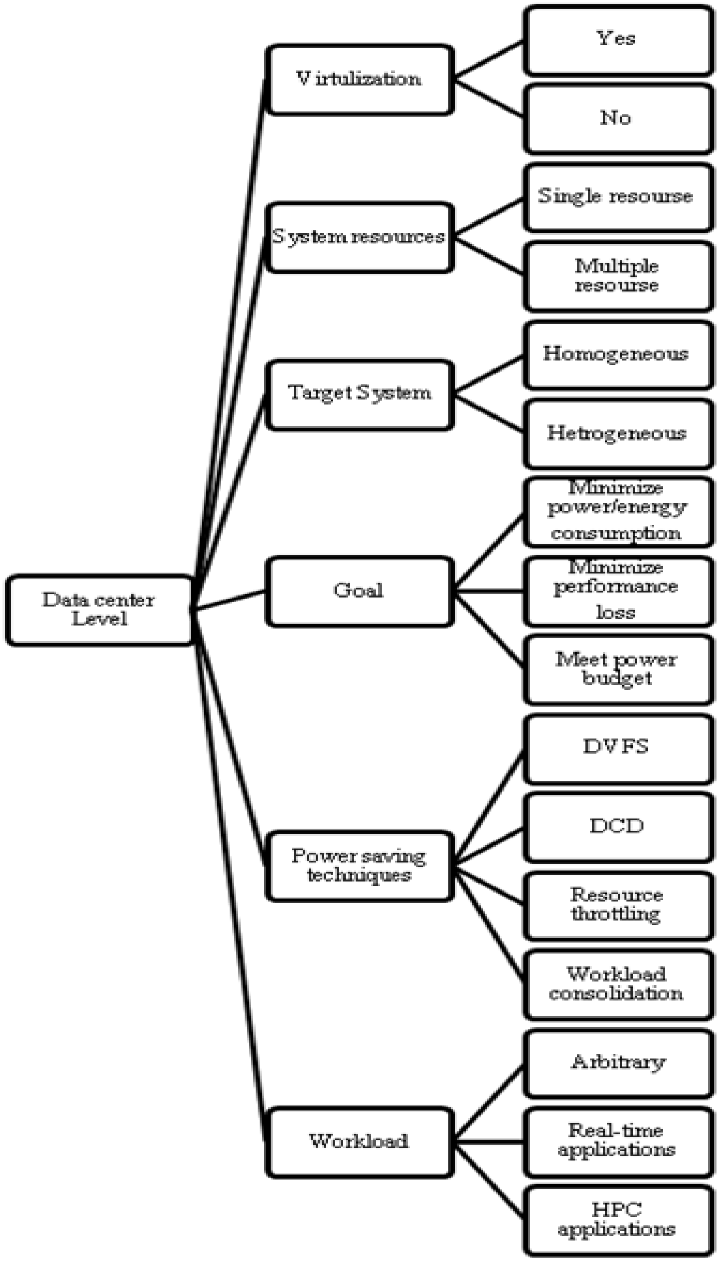

Figure.4 Data centre level taxonomy

7. RELATED WORK ON POWER MANAGEMENT

ENERGY-EFFICIENT SERVER CLUSTERS

Elnozahy et al. [5] have explored the problem of power-efficient resource management in a single-service environment for web applications with fixed SLA (response time) and automatic load-balancing running in a homogeneous cluster.
The motivation for the work is the reduction of operating costs and improvement of the error-proneness due to overheating. Two power management mechanisms that are applied switching physical nodes on and off (vary-on vary-off, VOVO) and DVFS of the CPU. The authors have proposed five policies for the resource management: independent voltage scaling (IVS), coordinated voltage scaling (CVS), VOVO, combined policy (VOVO-IVS), and coordinated combined policy (VOVO-CVS). The last Mentioned policy is stated to be the most advanced and is provided with a detailed description and mathematical model for determining CPU frequency thresholds. The thresholds define when it is appropriate to turn on an additional physical node or turn off an idle node. The main idea of the policy is to estimate total CPU frequency required to provide the expected response time, determine the optimal number of physical nodes, and proportionally set the frequency for all the nodes.

PMAPPER: POWER AND MIGRATION COST AWARE APPLICATION PLACEMENT

Verma et al. [6] have investigated the problem of dynamic placement of applications in virtualized systems, while minimizing the power consumption and maintaining the SLA. To address the problem, the authors have proposed the pMapper application placement framework. It consists of three managers and an arbitrator, which coordinates their actions and makes allocation decisions. Performance Manager monitors the applications' behaviour and resizes VMs according to current resource requirements and the SLA. Power Manager is in charge of adjusting hardware power states and applying DVFS. Migration Manager issues instructions for live migration of VMs in order to consolidate the workload. Arbitrator has a global view of the system and makes decisions about new placements of VMs and determines which VMs and which nodes should be migrated to achieve this placement. The authors have formulated the problem as a continuous optimization: at each time frame, the VM placement should be optimized to minimize the power consumption and maximize the performance. They make several assumptions to solve the problem, which are justified by experimental studies. The authors have presented several algorithms to solve the defined problem. They have defined it as a bin packing problem with variable bin sizes and costs.. To solve the bin packing problem, first-fit decreasing algorithm (FFD) has been adapted to work for differently sized bins with item-dependent cost functions. The first phase of mPP solves the cost minimization problem, whereas the second phase solves the application placement problem. trade-off . A VM is chosen to be migrated only if the revenue due to the new placement exceeds the migration cost. The experimental results show that the approach allows saving about 25% of power relatively to the Static and Load Balanced Placement algorithms.

ENERGY EFFICIENT ALLOCATION OF VIRTUAL MACHINES IN CLOUD DATA CENTRES

Anton and Buyya et al. [7] have proposed and evaluated heuristics for dynamic reallocation of VMs to minimize energy consumption, while providing reliable QoS. The obtained results show that the technique of dynamic reallocation of VMs and switching off the idle servers brings substantial energy savings and is applicable to real-world Cloud data centres. For the future work, they have proposed to investigate the consideration of multiple system resource in reallocation decisions, such as network interface and disk storage, as these resources also significantly contribute to the overall energy consumption. Other interesting directions for the future work suggested is investigation of setting the utilization thresholds dynamically according to a current set of VMs allocated to a host, leveraging multi-core CPU architectures, and decentralization of the optimization algorithms to improve scalability and fault tolerance. Besides the reduction of operational and establishment costs, the work has social significance as it decreases carbon dioxide footprints and energy consumption by modern IT infrastructures.

OPTIMAL POWER MANAGEMENT FOR SERVER FARM TO SUPPORT GREEN COMPUTING

Niyato, Chaisiri, & Sung et al. [8] has presented an approach which aims to contribute to the energy saving problem for data centers. Here the authors argue that efforts made previously were based on heuristic methods in which the optimal performance and power consumption cannot be guaranteed. So, they introduce a mechanism which works in two different sections of distributed data centers. First, each data center works along with an optimal power management module to make decisions about server mode switching to minimize the power consumption (turning on/off servers). Additionally, a module named job broker makes decisions on user's assignment to a specific data center with the aim of minimizing the total cost, which is composed of network and power consumption cost. Optimal performance levels are pursued trough turning on servers in advance to reduce the workloads' waiting time. The decision on how many servers should be reactivated is obtained by formulating and solving the constrained Markov decision process (CMDP). Additionally, optimizations at the job broker look for the best

workload placement within the different data centers trying to avoid job migrations among them that -as is mentioned by the authors- could lead to non-negligible system performance degradation. This approach considers the allocation of only one job per server. When the job is finished, the server sends a message to the scheduler indicating its status. Then the scheduler can assign a new job or deactivate it. This, in addition to the characteristic of awaking servers in advance, could be very beneficial to performance in scenarios where all the workloads had high computing demands. However, the energy savings in real cloud scenarios could be seriously affected because of the heterogeneity of workloads and the lack of mechanisms for handling heterogeneous hardware infrastructure. The allocation of jobs with low resource demands in complete servers could represent a serious resource waste problem.

PERFORMANCE EVALUATION OF A GREEN SCHEDULING ALGORITHM FOR ENERGY SAVINGS IN CLOUD COMPUTING

Duy, et al. [9] presents an approach which aims to contribute in the saving energy problem by allocating VMs to the least number of turned on servers. The difference remarked in this work with respect to others that also try to reduce the energy Consumption using "*dynamic servers' pool resizing*" is the introduction of an algorithm integrating a neural network predictor for optimizing server power consumption and reducing the performance impact in cloud computing environments. This neural network is used to anticipate the future load demand on servers by considering the historical demand with the aim of reducing the turning on/off frequency, and the resulting overhead which could lead to serious performance degradation. In this paper the authors describe the neural network, how it is composed, the training process, and how it works along with the "green" algorithm aiming to reduce the performance degradation.

POWER-AWARE SCHEDULING OF VIRTUAL MACHINES IN DVFS-ENABLED CLUSTERS

Von Laszewski, Wang, Younge, & He et al. [10] have described a scheduling mechanism which aims to reduce the power consumption in virtualized clustered environments by dynamically reducing processor speeds. The mechanism presented is composed of three algorithms that work together in order to allocate workloads in a virtualized cluster based on the required and available processor speed in the underlying physical nodes. The algorithms continuously monitor the VMs' status to adjust the processor speed on each node, reducing the power consumption. To achieve that, this approach uses profiles describing the available and maximum processor speed for each server in the cluster.

GREENCLOUD: A NEW ARCHITECTURE FOR GREEN DATA CENTER

Liu et al. [11] has described to reduce the power consumption in data centers by reducing the number of turned on servers. The authors present an architecture composed of some components such as monitoring services, a migration manager, the managed environment, and the front end that provides information to users. In this paper they are mainly focused in describing the live migration algorithm which search optimal placement of virtual machines, minimizing the total cost; being the cost in this paper calculated considering physical machine cost, the virtual machine status and the virtual machine migration cost.

8. CONCLUSION

Energy efficiency has emerged as one of the most important design requirements for modern computing systems, ranging from single servers to data centers and Clouds, as they continue to consume enormous amounts of electrical power. Apart from high operating costs incurred by computing resources, this leads to significant emissions of CO_2 into the environment. It is unacceptable in the age of climate change and global warming. To facilitate further developments in the area, it is essential to survey and review the existing body of knowledge. Therefore, in this paper, we have studied and classified various ways to achieve the power and energy efficiency in computing systems. Recent research advancements have been discussed and categorized across the virtualization and data centre levels. It has been shown that intelligent management of computing resources can lead to a significant reduction of the energy consumption by a system, while still meeting QoS and SLA.

REFERENCES

[1]Brad Nisbet- Monetizing the Cloud: XaaS Opportunities for Service Providers –Framingham, USAMarch 2011

[2]Anton Beloglazov and Rajkumar Buyya- Energy Efficient Resource Management in Virtualized Cloud Data Centers2011-2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing

[3]Cloud Computing Technology and Science (CloudCom), 2011 IEEE Third International Conference - Guazzone, M. Dept. of Comput. Sci., Univ. of Piemonte Orientale, Alessandria, Italy - Nov. 29 2011-Dec. 1 2011

[4]Nguyen Quang Hung, Nam Thoai, Nguyen Thanh Son *Ho* Performance Constraint And Power-Aware Allocation For User Request in Virtual Computing Lab , Ho Chi Minh City University of Technology, Vietnam August 12, 2011

[5]E. Elnozahy, M. Kistler, R. Rajamony ,Energy-efficient server clusters, Power Aware Computing System, 2325 (2003) 179–197.

[6]A. Verma, P. Ahuja, A. Neogi, pMapper: power and migration cost aware application placement in virtualized systems, presented at 9th ACM/IFIP/USENIX International Conference on Middleware, Springer-Verlag, New York, 2008, pp. 243–264.

[7]Anton Beloglazov and Rajkumar Buyya, Energy Efficient Allocation of Virtual Machines in Cloud Data Centers,2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing

[8]Niyato, D., Chaisiri, S., & Sung, L. B, Optimal Power Management for Server Farm to Support Green Computing. Paper presented at the 9th IEEE/ACM-2009 International Symposium on Cluster Computing and the Grid

[9]Duy, T. V. T., Sato, Y., & Inoguchi, Y. (2010). Performance evaluation of a Green Scheduling Algorithm for energy savings in Cloud computing. Presented at the IEEE International Symposium on Parallel & Distributed Processing, Workshops and Phd Forum (IPDPSW)

[10]Von Laszewski, G., Wang, L., Younge, A. J., & He, X. (2009). Power-Aware Scheduling of Virtual Machines in DVFS-enabled Clusters. Presented at the Proc. of IEEE International Conference on Cluster Computing 2009

[11]Liu, L., Wang, H., Liu, X., Jin, X., He, W. B., Wang, Q. B., GreenCloud: a new architecture for green data center. Paper presented at the Proc. of the sixth International Conference on Autonomic Computing Barcelona -2009, Spain

