International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IV, April 2018- Available at www.ijraset.com

Load balancing Technique for Energy-Energy Efficiency in Cloud Computing

Rani Danavath[1], Dr. V. B. Narsimha[2], Srinu Naik Dhanavath[3]
[1, 2, 3] *Osmania University*

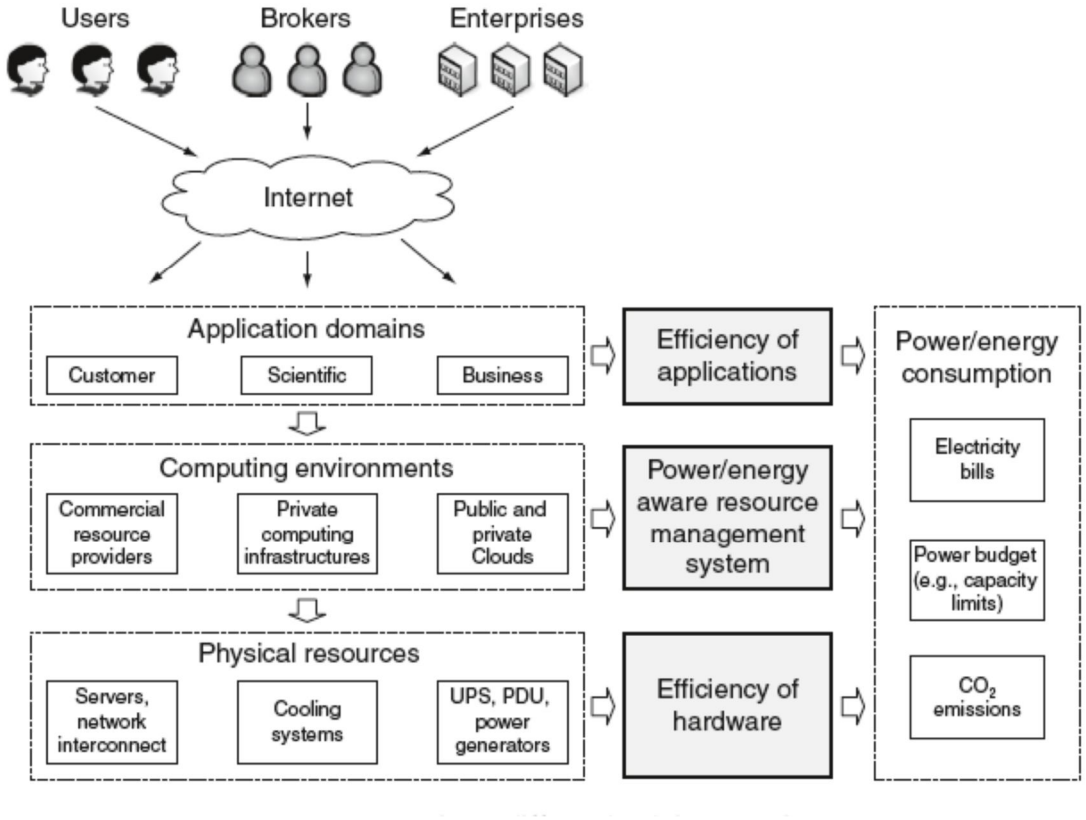
Abstract: ***Cloud computing is emerging as a new paradigm of large scale distributed computing. It is conceptually a distributed system where resources will be computing resources distributed through network and services pooled together and is provided to the users on pay-as-needed basis. Load balancing is one of the main challenges in cloud computing which is required to distribute the dynamic work load across multiple nodes to ensure that no single node is overloaded. It helps in optimal utilization of resources and hence its enhancing the performance of the system. The goal of the load balancing is minimising the resource consumption and carbon emission rate is the direct need of cloud computing. This determined the need of new metrics energy consumption and carbon emission for energy – efficiency load balancing techniques in cloud computing. Existing load balancing techniques mainly focuses on reducing overhead, services, response time and improving performance etc facts.***
In this paper we introduced a Technique .., but none of the techniques has considered the energy consumption and carbon emission that will go towards Energy – Efficiency. So this Energy – Efficiency load balancing technique can be used to improve the performance of cloud computing by balancing the workload across all the nodes in the cloud with the minimum resource utilization in turn the reducing energy consumption and carbon emission to an extent which will help to achieve green computing.
Keywords: cloud computing, Distributed computing, Energy efficiency, Green computing, Load balancing, Energy consumption and Carbon emission.

I. INTRODUCTION

Energy – Efficiency load balancing technique can improve the performance of the cloud computing. The energy consumption is not only determined by the efficiency of the physical resources, but it is also dependent on the resource management system deployed in the infrastructure and efficiency of applications running in the system. In this inter connection of the energy consumption and different levels of computing system can be seen from figure 1. Energy – efficiency impacts end users in terms of resource usage costs, which are typically determined by the Total Cost of Ownership (TCO) incurred by a resource provider. Higher power consumption results not only in boosted electricity bills, but also in additional requirements to a cooling system and power delivery infrastructure. i.e. Uninterruptable Power Slips (UPS), power Distribution Units (PDU), etc.., with the growth of computer components density the cooling system problem becomes crucial, as more heat has to be dissipated for a secure metor. The problem is especially important for 1 U and blade servers, These form factors are the most difficult to cool because of high density of the components ,ans thus lack of space for the air flow. Blade servers give the advantage of more computational power in less rock space.

For emanple—60 blade servers can be installed into a slandered 42v rock. however, such system requires more than 4,000w to supply the resources and cooling system compared to the same filled by 1 U server consumption tends to limit further performance improvements due to constraints of power distribution facilitates. For example, to power a server rack in typical data center, it is necessary to provide about 60Amps. Even if the cooling problem can be addressed for the future systems, it is likely that delivering current in such data centers will reach the power delivery limits.

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 (co2) emissions caused by high energy consumption. Therefore, the reduction of power and energy consumptions has become a first order objective in the design of modern computing systems. The roots of Energy – Efficient computing or Green It, practices can be traced back to 1992. When U.S environmental protection Agency launched energy star, evolunatory labelling problem whnic is designed to identify and promote Energy – Efficient products in order to reduce the Green House Gas Emission.



Energy consumption at different levels in computing systems.

There are a number of industry initiatives aiming at the development of standardized methods and techniques for reduction of the energy consumption in computer environment. They include climate savers computing initiative (CSCI), Green computing Impact Organization inc..,(GCIO), Green electronics council, The Green Grid, International professional practice partnerships (IP3), with membership of companies such as AMD, Dell, HP, IBM, Intel, Microsoft, Sun Micro systems and Vm ware.

A. Problem of Energy Efficiency Consumption

The energy consumption by computing facilities raises various monitory, environmental and system performance concerns. The scope of Energy – Efficiency design is not limited to main computing components (Eg: Processor, Storage devices, and visualization facilities), but it can expand into a much larger range of resources associated with computing facilities including auxiliary equipment's, water used for cooling and even physical floor space that these resources occupy.

While recent advances in hardware technologies including low power processors, solid state drivers and energy – Efficient monitors have alleviated the energy consumption issue to a certain degree, a series of software approaches have significantly contributed to the improvement of energy efficiency. These two approaches (hardware and software) should be seen as complementary rather than competitive user awareness is another non – negligible factor that should be taken into account when discussing Green IT. User awareness and behaviour in general considerably effect computing workload and resource usage pattern, this is in turn has a direct relationship with the energy consumption of not only core computing resource, but also auxiliary equipment, such as cooling air conditioning system.

For example a computer program developed without playing much attention to its energy efficiency may lead to excessive energy consumption and it may contribute to more heat emission resulting in increases in the energy consumption for cooling.

Traditionally, power and energy efficient resource management techniques have been applied to mobile devices, it was dictated by the fact that such devices are usually battery – powered and it is essential to consider power and energy management to improve their lifetime. However, due to continue growth of power and energy consumption by server and data centres, the focus of power and energy management techniques has been switched to these systems.

The energy consumption over a period of time. Therefore, the actual energy consumption by a data centre does not affect the cost of the infrastructure, On other hand, it is reflected in the electricity cost consumed by the system during the period of operation, which

is the main component of a data centres operating costs. Furthermore, in most data centres 50% of consumed energy never reaches the computing resources, it is consumed by the cooling facilities or dissipated in conversions within the UPS and PDU systems with the current tendency of continuously growing energy consumption and operating costs exceed the cost of computing resources , themselves in few years can be reached soon. Therefore, it is circuital to develop and apply Energy – Efficient resource management strategies in data centres.

Except for high operating costs, another problem caused by growing energy consumption is high Carbon Dioxide (Co2) emissions, which contribute to the global warming.

According to the estimation by the U.S Environmental protection Agency (EPA) of annual Co2 emissions from 42.8 million metrics tons (MMTCO2) in 2007 to 67.9 MMTCO2 in 2011. Intense media coverage has raised the awareness of people around the climate change and Green House Effect. More and more customers start to consider the "green" aspect in selecting products and services. Besides the environmental concern, business have begun to face risks caused by being non-environmentally friendly. Reduction of co2 footprints is an important problem that has to be addressed in order to facilitate further advancements in computing system.

B. Models of power and Energy

Power and Energy can be defined in terms of work that a system performs.

$$P = W/T$$

$$E = P.T$$

1) Static and dynamic power consumption models.: The main power consumption in complementory Metal – Oxide – Semiconductor (CMOS) circutes comprises static and dynamic power consumption models.

2) Sources of power consumption model

3) Modeling power consumption model.

C. Power and Energy consumption levels

1) Hardware and Firmware level

2) Operating system level.

3) Virtualization level.

4) Data center level.

In this paper we introduced a technique for Energy – Efficiency at the Virtualization level.

D. System architecture

In this work the underlying infrastructure is represented by a large-scale Cloud data center comprising *n* heterogeneous physical nodes. Each node has a CPU, which can be multicore, with performance defined in Millions Instructions Per Second (MIPS). Besides that, a node is characterized by the amount of RAM and network bandwidth. Users submit requests for provisioning of *m* heterogeneous VMs with resource requirements defined in MIPS, amount of RAM and network bandwidth. SLA violation occurs when a VM cannot get the requested amount of resource, which may happen due to VM consolidation. The software system architecture is tiered comprising a dispatcher, global and local managers. The local managers reside on each physical node as a part of a Virtual Machine Monitor (VMM). They are responsible for observing current utilization of the node's resources and its thermal state. The local managers choose VMs that have to be migrated to another node in the following cases:

1) The utilization of some resource is close to 100% that creates a risk of SLA violation.

2) The utilization of resources is low, therefore, all the VMs should be reallocated to another node and the idle node should be turned off.

3) A VM has intensive network communication with another VM allocated to a different physical host.

4) The temperature exceeds some limit and VMs have to be migrated in order to reduce load on the cooling system and allow the node to cool down naturally.

The local managers send to the global managers the information about the utilization of resources and VMs chosen to migrate. Besides that, they issue commands for VM resizing, application of DVFS and turning on / off idle nodes. Each global manager is attached to a set of nodes

and processes data obtained from their local managers. The global managers continuously apply distributed version of a heuristic for semi-online multidimensional bin-packing, where bins represent physical nodes and items are VMs that have to be allocated. The decentralization removes

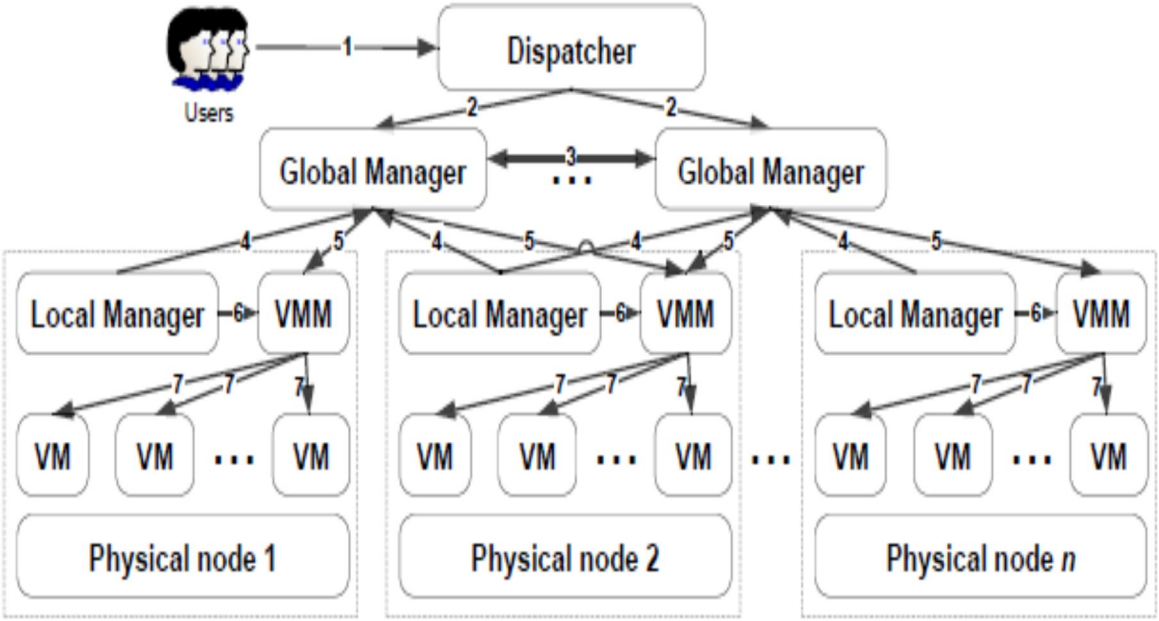


Figure 1. The system architecture

a Single Point of Failure (SPF) and improves scalability. Each dimension of an item represents the utilization of a particular resource. After obtaining allocation decision, the global managers issue commands for live migration of VMs. As shown in Figure 1, the system operation consists of the following steps:

a) New requests for VM provisioning. Users submit requests for provisioning of VMs.
b) Dispatching requests for VM provisioning. The dispatcher distributes requests among global managers.
c) Intercommunication between global managers. The global managers exchange information about utilization of resources and VMs that have to be allocated.
d) Data about utilization of resources and VMs chosen to migrate. The local managers propagate information about resource utilization and VMs chosen to migrate to the global managers.
e) Migration commands. The global managers issue VM migration commands in order to optimize current allocation.
f) Commands for VM resizing and adjusting of power states. The local managers monitor their host nodes and issue commands for VM resizing and changes in power states of nodes.
g) VM resizing, scheduling and migration actions. According to the received commands, VMM performs actual resizing and migration of VMs as well as resource scheduling.

E. Evoluation

As the proposed system is targeted on a large-scale Cloud data center, it is necessary to conduct large-scale experiments to evaluate the algorithms. However, it is difficult to run large-scale experiments on a real-world infrastructure, especially when the experiments have to be repeated for
different policies with the same conditions [18]. Therefore, simulation has been chosen as a way to evaluate the proposed heuristics. We have chosen CloudSim toolkit [18] as a simulation framework, as it is built for simulation of Cloud computing environments. In comparison to alternative simulation

toolkits (e.g. SimGrid, GangSim), CloudSim supports modeling of on-demand virtualization enabled resource and application management. We have extended the framework in order to enable energy aware simulations as the core framework does not provide this capability. In addition, we have incorporated the abilities to account SLA violations and to simulate dynamic workloads that correspond to web applications and online services. The simulated data center consists of 100 heterogeneous physical nodes. Each node is modeled to have one CPU core with performance equivalent to 1000, 2000 or 3000 MIPS, 8 Gb of RAM and 1 TB of storage. Users submit requests

for provisioning of 290 heterogeneous VMs that fill the full capacity of the data center. For the borderline policies we simulated a Non Power Aware policy (NPA) and DVFS that adjusts the voltage and frequency of CPU according to current utilization. We simulated a Single Threshold policy (ST) and two-threshold policy aimed at Minimization of Migrations (MM). Besides that, the policies have been evaluated with different values of the thresholds.

Table I
SIMULATION RESULTS

Policy	Energy	SLA	Migr.	Avg. SLA
NPA	9.15 KWh	-	-	-
DVFS	4.40 KWh	-	-	-
ST 50%	2.03 KWh	5.41%	35 226	81%
ST 60%	1.50 KWh	9.04%	34 231	89%
MM 30-70%	1.48 KWh	1.11%	3 359	56%
MM 40-80%	1.27 KWh	2.75%	3 241	65%
MM 50-90%	1.14 KWh	6.69%	3 120	76%

The simulation results are presented in Table I. The results show that dynamic reallocation of VMs according to current utilization of CPU can bring higher energy savings comparing to static allocation policies. MM policy allows to achieve the best energy savings: by 83%, 66% and 23% less energy consumption relatively to NPA, DVFS and ST policies respectively with thresholds 30-70% and ensuring percentage of SLA violations of 1.1%; and by 87%, 74% and 43% with thresholds 50-90% and 6.7% of SLA violations. MM policy leads to more than 10 times fewer VM migrations than ST. The results show the flexibility of the algorithm, as the thresholds can be adjusted according to SLA requirements. Strict SLA (1.11%) allow achievement of the energy consumption of 1.48 KWh. However, if SLA are relaxed (6.69%), the energy consumption is further reduced to 1.14 KWh.

II. CONCLUSION

In this paper have presented a decentralized architecture of the energy aware resource management system for Cloud data centers. We have defined the problem of minimizing

the energy consumption while meeting QoS requirements and stated the requirements for VM allocation policies. Moreover, we have proposed three stages of continuous optimization of VM placement and presented heuristics for a simplified version of the first stage. The heuristics have

been evaluated by simulation using the extended CloudSim toolkit. One of the heuristics leads to significant reduction of the energy consumption by a Cloud data center – by 83% in comparison to a non power aware system and by 66% in comparison to a system that applies only DVFS

technique but does not adapt allocation of VMs in run-time. Moreover, MM policy enables flexible adjustment of SLA by setting appropriate values of the utilization thresholds: SLA can be relaxed leading to further improvement of energy consumption. The policy supports heterogeneity of both the

hardware and VMs and does not require any knowledge about particular applications running on the VMs. The policy is independent of the workload type.

REFERENCES

[1] Anton Beloglazov, Rajkumar Byya, Young choon Lee, and Albert Zomany " A Taxonomy and survey of efficient data centers and Cloud Computing Systems."
[2] Anton Beloglazov* and Rajkumar Buyya Cloud Computing and Distributed Systems (CLOUDS) Laboratory" Energy Efficient Resource Management in Virtualized Cloud Data Centers"

[3] P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R. Neugebauer, I. Pratt, and A. Warfield, "Xen and the art of virtualization," in Proceedings of the 19th ACM symposium on Operating systems principles, 2003, p. 177.
[4] R. Buyya, C. S. Yeo, and S. Venugopal, "Market-oriented cloud computing: Vision, hype, and reality for delivering it services as computing utilities," in Proceedings of the 10th IEEE International Conference on High Performance Computing and Communications (HPCC'08). IEEE CS Press, Los Alamitos, CA, USA, 2008
[5] R. Neugebauer and D. McAuley, "Energy is just another resource: Energy accounting and energy pricing in the nemesis OS," in Proceedings of the 8th IEEE Workshop on Hot Topics in Operating Systems, 2001, pp. 59–64.
[6] H. Zeng, C. S. Ellis, A. R. Lebeck, and A. Vahdat, "ECOSystem: managing energy as a first class operating system resource," ACM SIGPLAN Notices, vol. 37, no. 10, p. 132, 2002.
[7] E. Pinheiro, R. Bianchini, E. V. Carrera, and T. Heath, "Load balancing and unbalancing for power and performance in cluster-based systems," in Workshop on Compilers and Operating Systems for Low Power,
[8] G.E. Moore, Cramming more components onto integrated circuits, Proc. IEEE 2001, pp. 182–195.86 (1) (1998) 82–85.
[9] J.G. Koomey, Estimating Total Power Consumption by Servers in the US and the World, Analytics Press, Oakland, CA, 2007.
[10] L. Barroso, The Price of Performance, ACM Press, Queue, 2005, vol. 3 (7), p. 53.
[11] R. Brown, E. Masanet, B. Nordman, B. Tschudi, A. Shehabi, J. Stanley, et al., Report to Congress on Server and Data Center Energy Efficiency: Public Law 109–431, Lawrence Berkeley National Laboratory, Berkeley, CA, USA, 2008.
[12] L. Minas, B. Ellison, Energy Efficiency for Information Technology: How to Reduce Power Consumption in Servers and Data Centers, Intel Press, USA, 2009.