International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 22 Issue 1 – MAY 2016.

ENERGY OPTIMIZATION USING VIRTUAL MACHINE IN CLOUD

Karthigai Raja. T [1] , Suma Sira Jacob [2]

[1] PG Scholar, Department of Computer Science and Engineering, Christian College of Engineering and Technology, Oddanchatram.Tamil Nadu. India.
[2] M.Tech (Ph.d), Assistant Professor, Department of Computer Science and Engineering, Christian College of Engineering and Technology, Oddanchatram.Tamil Nadu. India

***Abstract*--The energy consumption of computer and communication systems does not scale linearly with the workload. A system uses a significant amount of energy even when idle or lightly loaded. A widely reported solution to resource management in large data centers is to concentrate the load on a subset of servers and, whenever possible, switch the rest of the servers to one of the possible sleep states. In this paper we introduce an energy-aware operation model used for load balancing and application scaling on a cloud. We also propose visual representation of a process which is being running on each virtual machine. To make workload synchronous the process are added to queue. They will be served by virtual machines one by one by which virtual machines is in sleep stage. Another advantage is each virtual machines status information is updated and monitored by cloud.**

Index terms: - *Load Balancing, Application Scaling, Idle Servers, Server Consolidation, Energy Proportional Systems*.

I. INTRODUCTION

The concept of “load balancing" dates back to the time when the first distributed computing systems were implemented. It means exactly what the name implies, to evenly distribute the workload to a set of servers to maximize the throughput, minimize the response time, and increase the system resilience to faults by avoiding overloading the systems.

Scaling is the process of allocating additional resources to a cloud application in response to a request consistent with the SLA. We distinguish two scaling modes, horizontal and vertical scaling. Horizontal scaling is the most common mode of scaling on a cloud; it is done by increasing the number of Virtual Machines (VMs) when the load of applications increases and reducing this number when the load decreases.

Load balancing is critical for this mode of operation. Vertical scaling keeps the number of VMs of an application constant, but increases the amount of resources allocated to each one of them.

This can be done either by migrating the VMs to more powerful servers or by keeping the VMs on the same servers, but increasing their share of the server capacity. The first alternative involves additional overhead; the VM is stopped, a snapshot is taken, the file is migrated to a more powerful server, and the VM is restarted at the new site.

II. SYSTEM MODEL

In this section we consider the existing system design and the proposed system.

2.1 EXISTING SYSTEM

The existing design approaches an important strategy for energy reduction is concentrating the load on a subset of servers and, whenever possible, switching the rest of them to a state with low energy consumption. This observation implies that the traditional concept of load balancing in a large-scale system could be reformulated as follows: distribute evenly the workload to the smallest set of servers operating at optimal or near-optimal energy levels, while observing the Service Level Agreement (SLA) between the CSP and a cloud user. An optimal energy level is one when the performance per Watt of power is maximized.

In order to integrate business requirements and application level needs, in terms of Quality of Service (QoS), cloud service provisioning is regulated by Service Level Agreements (SLAs): contracts between clients and providers that express the price for a service, the QoS levels required during the service provisioning, and the penalties associated with the SLA violations.

In such a context, performance evaluation plays a key role allowing system managers to evaluate the effects of different resource management strategies on the data center functioning and to predict the corresponding costs/benefits.

Problem Identified:

- On-the-field experiments are mainly focused on the offered QoS, they are based on a black box approach that makes difficult to correlate obtained data to the internal resource management strategies implemented by the system provider.
- Simulation does not allow to conduct comprehensive analyses of the system performance due to the great number of parameters that have to be investigated.

2.2 PROPOSED SYSTEM

There are three primary contributions of this paper: a new model of cloud servers that is based on different operating regimes with various degrees of \energy efficiency" (processing power versus energy consumption);a novel algorithm that performs load balancing and application scaling to maximize the number of servers operating in the energy-optimal regime; and analysis and comparison of techniques for load balancing and application scaling using three differentlysized clusters and two different average load profiles. The objective of the algorithms is to ensure that the largest possible number of active servers operate within the boundaries of their respective optimal operating regime.

The actions implementing this policy are: (a) migrate VMs from a server operating in the undesirable-low regime and then switch the server to a sleep state; (b) switch an idle server to a sleep state and reactivate servers in a sleep state when the cluster load increases; (c) migrate the VMs from an overloaded server, a server operating in the undesirable-high regime with applications predicted to increase their demands for computing in the next reallocation cycles.

Benefits:

- After load balancing, the number of servers in the optimal regime increases from 0 to about 60% and a fair number of servers are switched to the sleep state.
- There is a balance between computational efficiency and SLA violations; the algorithm can be tuned to maximize computational efficiency or to minimize SLA violations according to the type of workload and the system management policies.

III. DESIGN CONSTRUCTION

This Section consists of the following module design are to be explained in this section.

3.1 SYSTEM ARCHITECTURE

The model introduced in this section assumes a clustered organization of the cloud infrastructure and targets primarily The IaaS cloud delivery model represented by Amazon Web Services (AWS).

AWS supports a limited number of instance families, including M3 (general purpose), C3 (compute optimized), R3 (memory optimized), I2 (storage optimized), G2 (GPU) and so on.

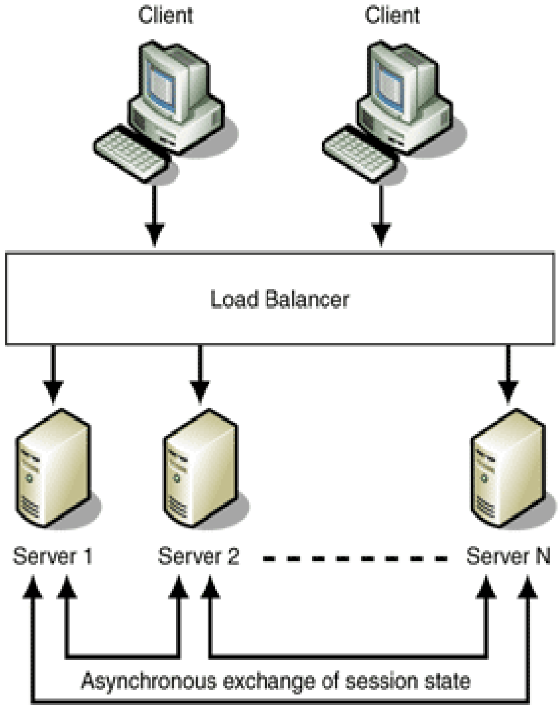


Fig-1: Generation of System Architecture Model

3.1.1 Clustered organization:

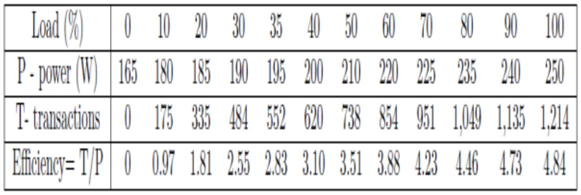
A cluster C has a leader, a system which maintains relatively accurate information about the free capacity of individual servers in the cluster and communicates with the leaders of the other clusters for the implementation of global resource management policies. The leader could consist of a multi-system configuration to guarantee a fast response time and to support fault-tolerance.

3.1.2 System and application level resource management:

The model is based on a two-level decision making process, one at the system and one at the application level. The scheduler of the virtual Machine Monitor (VMM) of a server interacts with the Server Application Manager (SAM) component of the VMM to ensure that the QoS requirements of the application are satisfied [1]. SAM gathers information from individual application managers of the VMs running on the server.

3.1.3 Measuring server energy efficiency:

A recent benchmark [2] compares the energy efficiency of typical business applications running on a Java platform. From Table 1 we see that the energy efficiency is nearly linear.



Load (%)	0	10	20	30	35	40	50	60	70	80	90	100
P - power (W)	165	180	185	190	195	200	210	220	225	235	240	250
T- transactions	0	175	335	484	552	620	738	854	951	1,049	1,135	1,214
Efficiency= T/P	0	0.97	1.81	2.55	2.83	3.10	3.51	3.88	4.23	4.46	4.73	4.84

3.2 ENERGY-AWARE SCALING ALGORITHMS:

The objective of the algorithms is to ensure that the largest possible number of active servers operate within the boundaries of their respective optimal operating regime.

The actions implementing this policy are:

(a) Migrate VMs from a server operating in the undesirable-low regime and then switch the server to a sleep state.
(b) Switch an idle server to a sleep state and reactivate servers in a sleep state when the cluster load increases.
(c) Migrate the VMs from an overloaded server, a server operating in the undesirable-high regime with applications predicted to increase their demands for computing in the next reallocation cycles.

3.2.1 Scaling decisions:

The Server Application Manager SAM is a component of the Virtual Machine Monitor (VMM) of a server S.

3.2.2 Cluster management:

The leader of a cluster maintains several control structures:

3.2.2.1. Authentication

Now peoples are chosen cloud to use storage more secure. So the authentication is required for protect data security and user privacy. Here the User validated whether he is valid site user or not. To join to our cloud the customer must register and he will get login credentials for authentication. Next time he will login by use this credential information.

3.2.2.2. Customer Request

Each customer can make a request for hosting their data and update it by the way of making request. Customer request are effectively handled by the controller which is the part of vm shedular. The Request may be uploading data, deleting data modifying data in a cloud environment.

3.2.2.3. VM Simulator

In this module we developed the simulator which is being useful to make the interaction between administrator and vm hardware. Each Virtual Machine simulation state is monitored by this simulator and their performance will be measured up to date. The simulation are including memory they are using, power consumption, energy using and also their status whether it is in sleep or running.

3.2.2.4. Process Queue.

To handle the subsequent process from customer in various domains all must be maintained in stack or list. Here we uses the Queue (First In First Out) for handling that requests and they are served by one by one by virtual machine which is in sleep stage . After completion of each process the vm is kept into sleep stage from running by the controller to save energy.

3.2.2.5. Feedback

Customers feedbacks are collected for future implementation of that private cloud and also for rectifying their complaints and for provide solutions.

IV. PERFORMANCE EVALUATION

The effect of average server load on the distribution of the servers in the five operating regimes. R1, R2, R3, R4 and R5, before and after energy optimization and load balancing.

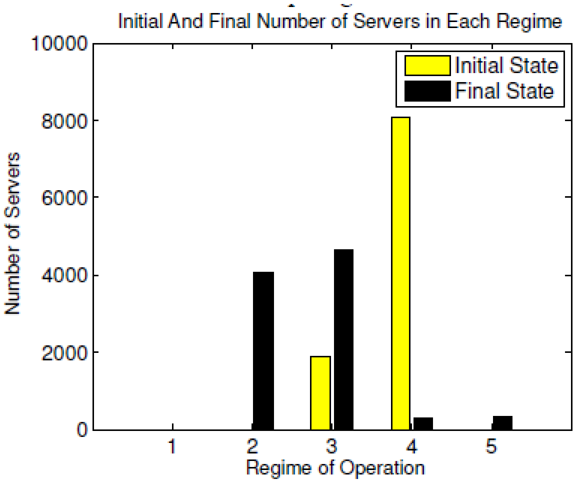
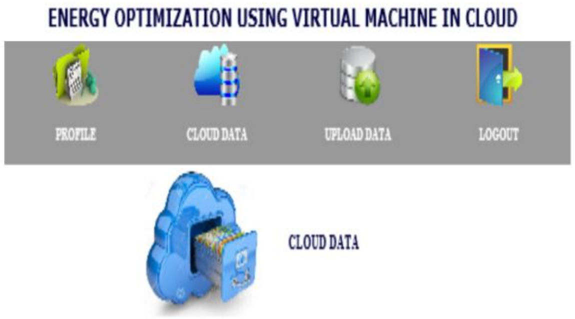
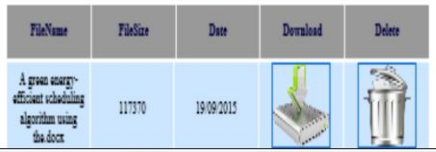


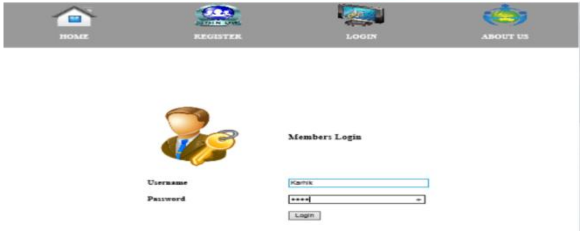
Fig-2: Cluster size 104. Average load: 70%

V. SIMULATION RESULT





FileName	FileSize	Date	Download	Delete
A green energy-efficient scheduling algorithm using the.docx	117370	19/09/2015		

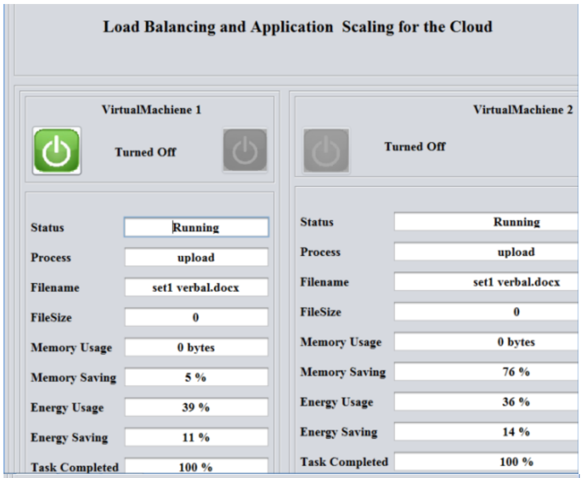


Member: Login

Username Karhik

Password ****

Login



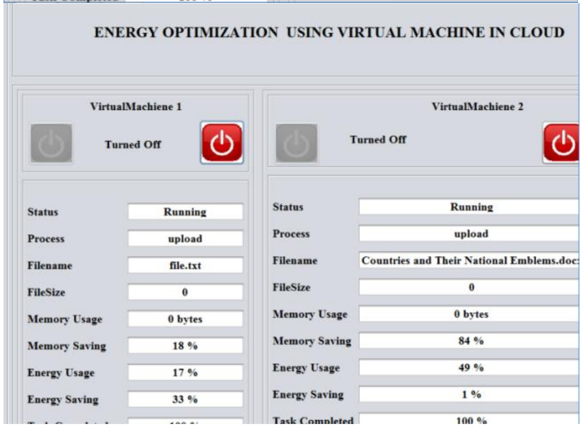
Load Balancing and Application Scaling for the Cloud

VirtualMachiene 1 — Turned Off

Status	Running
Process	upload
Filename	set1 verbal.docx
FileSize	0
Memory Usage	0 bytes
Memory Saving	5 %
Energy Usage	39 %
Energy Saving	11 %
Task Completed	100 %

VirtualMachiene 2 — Turned Off

Status	Running
Process	upload
Filename	set1 verbal.docx
FileSize	0
Memory Usage	0 bytes
Memory Saving	76 %
Energy Usage	36 %
Energy Saving	14 %
Task Completed	100 %



ENERGY OPTIMIZATION USING VIRTUAL MACHINE IN CLOUD

VirtualMachiene 1 — Turned Off

Status	Running
Process	upload
Filename	file.txt
FileSize	0
Memory Usage	0 bytes
Memory Saving	18 %
Energy Usage	17 %
Energy Saving	33 %

VirtualMachiene 2 — Turned Off

Status	Running
Process	upload
Filename	Countries and Their National Emblems.doc:
FileSize	0
Memory Usage	0 bytes
Memory Saving	84 %
Energy Usage	49 %
Energy Saving	1 %
Task Completed	100 %

VI. CONCLUSION

The realization that power consumption of cloud computing centers is significant and is expected to increase substantially in the future motivates the interest of the research community in energy-aware resource management and application placement policies and the mechanisms to enforce these policies. A quantitative evaluation of an optimization algorithm or an architectural enhancement is a rather intricate and time consuming process; several benchmarks and system configurations are used to gather the data necessary to guide future developments.

Future Work: Our future work will evaluate the overhead and the limitations of the algorithm proposed in this paper; it will also include the implementation of a Server Application Manager and the evaluation of the overhead for the algorithm proposed in this paper. The algorithm will be incorporated in the self management policies.

REFERENCE

[1].D. Ardagna, B. Panicucci, M. Trubian, and L. Zhang. “Energy-aware autonomic resource allocation in multi- tier virtualized environments.” IEEE Trans. on Services Computing, 5(1):2–19, 2012.

[2].L. A. Barroso and U. H¨ozle. “The case for energy- proportional computing.” IEEE Computer, 40(12):33– 37, 2007.

[3].A. Beloglazov, R. Buyya “Energy efficient resource man- agement in virtualized cloud data centers.” Proceedings of the 2010 10th IEEE/ACM - 2010.

[4].D. Gmach, J. Rolia, L. Cherkasova, G. Belrose, T. Tu- cricchi, and A. Kemper. “An integrated approach to re- source pool management: policies, efficiency, and quality metrics.”. 326–335, 2008.

[5]. Ashkan Paya and Dan C. Marinescu, “Energy-aware Load Balancing and Application Scaling for the Cloud Ecosystem”, IEEE Transactions on Cloud Computing-2015.