Volume 4 Issue III, March 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Survey on Efficient Control Algorithm in Cloud for Cost Elaboration

R.Vinoth[1], R. Karthik[2]
[1]Assistant Professor, [2]Final Year Student, Department Of Information Technology,
A.V.C. College of Engineering, Mannampandal, Mayiladuthurai, Nagapatinam District, Tamilnadu, India.

Abstract: Cloud computing is a new paradigm for delivering remote computing resources through a network. However, achieving an energy-efficiency control and simultaneously satisfying a performance guarantee have become critical issues for cloud providers. A cost function has been developed in which the costs of power consumption, system congestion and server start up are all taken into consideration.The effect of energy-efficiency controls on response times, operating modes and incurred costs are all demonstrated. Our objectives are to find the optimal service rate and mode-switching restriction, so as to minimize cost within a response time guarantee under varying arrival rates. Failure to consider the heterogeneity of both machines and workloads will lead to both sub-optimal energy-savings and long scheduling delays, due to incompatibility between workload requirements and the resources offered by the provisioned machines. An efficient control (EC) algorithm is first proposed for solving constrained optimization problems and making costs/performances tradeoffs in systems with different power-saving policies.

I. INTRODUCTION

Cloud computing is a new service model for sharing a pool of computing resources that can be rapidly accessed based on a converged infrastructure. In the past, an individual use or company can only use their own servers to manage application programs or store data. Nowadays, resources provided by cloud allow users to get on demand access with minimal management effort based on their needs. Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) are all existing service models. For example, Amazon web services is a well-known IaaS that lets users perform computations on the Elastic Compute Cloud (EC2). Google's App Engine and Sales force are public clouds for providing PaaS and SaaS, respectively. This vision of computing utilities based on a service provisioning model anticipated the massive transformation of the entire computing industry in the 21st century where by computing services will be readily available on demand, like other utility services available in today's society. First, burst arrivals may experience latency or be unable to access services. Second, there has a power consumption overhead caused by awakening servers from a power-off state too frequently. Third, the worst case is violating a service level agreement (SLA) due to the fact that shutting down servers may sacrifice quality of service. Green Cloud computing is envisioned to achieve not only efficient processing and utilization of computing infrastructure, but also minimize energy consumption.

II. RELATED WORK

Power savings in cloud systems have been extensively studied on various aspects in recent years, e.g., on the virtual machine (VM) side by migrating VMs, applying consolidation or allocation algorithms, and on the data center infrastructure side through resource allocations, energy managements, etc.

A. Power-Saving In Virtual Machine

Considered the problem of providing power budgeting support while dealing with many problems that arose when budgets virtualized systems.Their approach to VM-aware power budgeting used multiple distributed managers integrated into the virtual power management (VPM) framework.By investigated the potential performance overheads caused by server consolidation and lived migration of virtual machine technology.The potential performances overheads of server consolidation were evaluated.

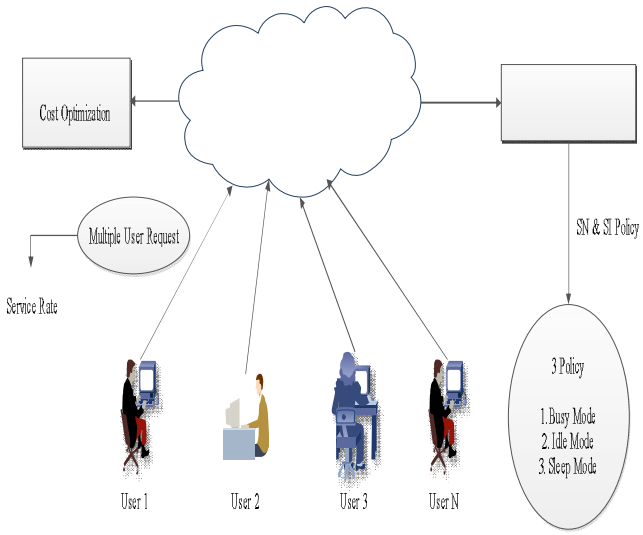
B. Power-Saving In Computing Infrastructure

The Datacenter Energy Management project was focused on modeling energy consumption in data centers, with a goal to optimize electricity consumption. Their project was focused on collecting data to define basic fuel consumption curves.A Heterogeneity-Aware Resource Monitoring and management system that was capable of performing dynamic capacity provisioning (DCP) in

Volume 4 Issue III, March 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

heterogeneous data Centers.

III. ARCHITECTURE

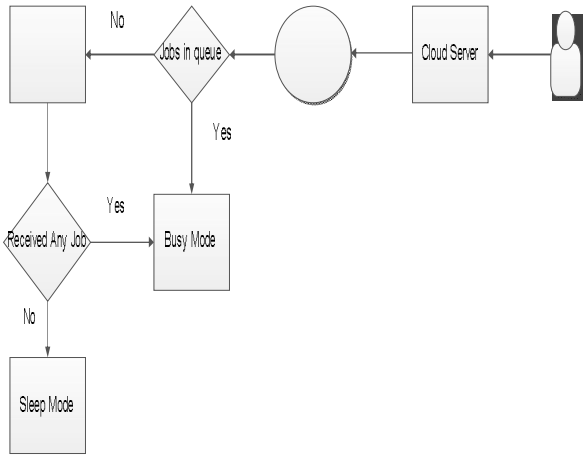


IV. POLICIES

They are 3 policies used
ISN policy
SN policy
SI policy

A. ISN Policy

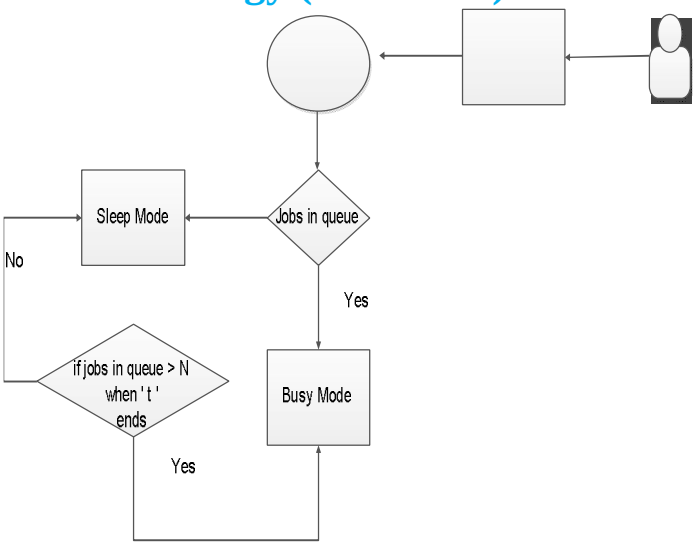
An energy-efficient control in a system with three operating modes m = {Busy, Idle, Sleep}, where a sleep mode would be responsible for saving power consumption.A server is allowed to stay in an idle mode for a short time when there has no job in the system, rather than switch abruptly into a sleep mode right away when the system becomes empty. An idle mode is the only operating mode that connects to a sleep mode.



B. SN Policy

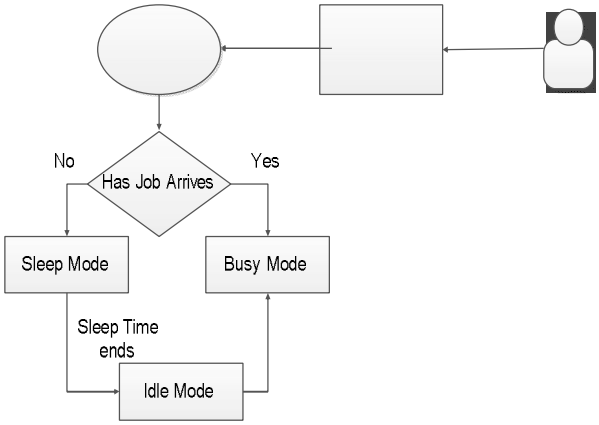
According to the switching process (directly to Sleep) and the energy-efficient control (N policy), we have called such an approach the “SN policy”. A server switches into a sleep mode immediately when no job is in the system. A server stays in a sleep mode if the number of jobs in the queue is less than the N value; otherwise, a server switches into a busy mode and begins to work.

Volume 4 Issue III, March 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)



C. SI Policy

A server switches into a sleep mode immediately instead of an idle mode when there has no job in the system. A server can stay in a sleep mode for a given time in an operation period. If there has no job arrival when a sleeping time expires, a server will enter into an idle mode.



V. MATHEMATICAL REPRESENTATIONS

A. Queuing Models

Systems applied with these power-saving policies follow the identical assumptions as follows. It is assumed that job request arrivals follow a Poisson process with parameter and they are served in order of their arrivals, that is, the queue discipline is the first come first served (FCFS).

P0n _ Probability that there have n jobs in the system when a server is in a busy mode;
P1 _ Probability that there has no job in the system when a server is in an idle mode;
P2n _ Probability that there have n jobs in the system when a server is in a sleep mode.

B. Optimization Problem Formulation

In general, a larger controlled N value can gain more power saving but result in excessive delay. Conversely, a smaller controlled N value can reduce delay times but lead to a shorter operational cycle. Therefore, the power consumption overhead due to server start up cannot be ignored.

P0 _ Probability that there has no job in the system and no stage begins working when a server is in a sleep mode;
P0k _ Probability that there have i jobs in the system and a job is processed at the stage k when a server is in a sleep mode, where i

=1; 2; . . .;N _ 1;
P1ij _ Probability that there have i jobs in the system and a job is processed at the stage j when a server is in a busy mode, where i =1; 2; . . . ; and j =1; 2; . . . ; k.

C. Performance Comparisons And The ECG Control Algorithm

To gain more insight into systems with different power-saving policies, experiments are conducted to illustrate the relationship between the mode-switching restriction and traffic-load intensity on power consumption cost and system congestion cost;
Examine the idle and sleep probability distributions under different service rates and
Compare response times and total operational costs with a typical system, where it doesn't have any energy-efficient control.

D. Algorithm

An efficient control (EC) algorithm is first proposed for solving constrained optimization problems and making costs/performances tradeoffs in systems with different power-saving policies.
Input:
1. An arrival rate
2. Upper bound of the server rate and the waiting buffer, denoted by mu and Nb.
3. Cost parameters [Co; C1; . . . ; C6].
4. A response time guarantee x.
5. System parameters {Qi;Qd;Qs} used by the ISN policy.
6. System parameter {k} used by the SN policy.
7. System parameters {Q;N = 1} used by the SI policy.
Output:
Step 1. For i = 1; i = u; i++
Set mi ->a current service rate;

Step 2. For j = 1; j = b; j++
Set Nj-> a current N parameter;

Step 3. Calculate the system utilization.
If the current test parameters satisfy the constraint of (i)0 <r < 1,
then
Calculate the response time;
Else
Return to step 1 and begin to test a next index;
End

Step 4. If the current test parameters satisfy the constraint of (ii) W <x,
then
Record the current joint values of (mi;Nj) and identify it as the approved joint parameters;
Else
Return to step 1 and begin to test a next index;
End

Step 5. When all the test parameters have been done, then
current set of the approved parameters;
Bring cost parameters into the objective function by using Eq. (6) and test all approved joint parameters;

Step 6. If the joint values of (mi;nj) can obtain the minimum cost value in all testing, then,Output (mi;Nj) and (mi;Nj)
Else
Return to step 5 and begin to test a next approved parameter.

End

VI. CONCLUSION

The growing crisis in power shortages has brought a concern in existing and future cloud system designs. To mitigate unnecessary idle power consumption, three power saving policies with different decision processes and mode switching controls are considered. Our proposed algorithm allows cloud providers to optimize the decision-making in service rate and mode-switching restriction, so as to minimize the operational cost without sacrificing a SLA constraint. As compared to a general policy, cost savings and response time improvement can be verified. we present Harmony, a heterogeneity-aware framework that dynamically adjusts the number of machines to strike a balance between energy savings and scheduling delay, while considering the reconfiguration cost.

REFERENCES

[1] Yi-Ju Chiang and Ching-Hsien"An Efficient Green Control Algorithm in Cloud Computing for Cost Optimization,"IEEE TRANSACTIONS ON CLOUD COMPUTING, VOL. 3, NO. 2, APRIL/JUNE 2015
[2] Y. C. Lee and A. Y. Zomaya, "Energy efficient utilization of resources in cloud computing systems," J. Supercomputer., vol. 60, no. 2, pp. 268–280, 2012.
[3] Amokrane,M. Zhani,R.Langar,R.Boutaba and G.Pujolle, "Greenhead: Virtual data center embedding across distributed infrastructures," IEEE Trans. Cloud Computer., vol. 1, no. 1, pp. 36– 49, Jan.–Jun. 2013
[4] Q. Zhang, M. Zhani, R. Boutaba, and J. Hellerstein, "Dynamic heterogenius aware resource provisioning in the cloud," IEEE Trans. Cloud Computer., vol. 2, no. 1, pp. 14–28, Jan.–Mar. 2014
[5] G. Wang and T. E. Ng, "The impact of virtualization on network performance of amazon ec2 data center," in Proc. IEEE Proc. INFOCOM, 2010, pp. 1–9.
[6] R. Ranjan, L. Zhao, X. Wu, A. Liu, A. Quiroz, and M. Parashar, "Peer-to-peer cloud provisioning: Service discovery and load-balancing," in Cloud Computing. London, U.K.: Springer, 2010, pp. 195–217.
[7] M. Guazzone, C. Anglano and M. Canonico, "Energy-efficient resource management for cloud computing infrastructures," in Proc. IEEE Int. Conf. Cloud Comput. Technol. Sci., 2011, pp. 424–431.
[8] D. A. Wu and H. Takagi, "M/G/1 queue with multiple working vacations," Perform. Eval., vol. 63, no. 7, pp. 654–681, 2006.
[9] M. Yadin and P. Naor, "Queueing systems with a removable service station," Operations Res., vol . 14, pp. 393–405, 1963.
[10] J. Song, T. Li, Z. Wang, and Z. Zhu, "Study on energy-consumption regularities of cloud computing systems by a novel evaluation model," Computing, vol. 95, no. 4, pp. 269–287, 2013

