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

COST OPTIMIZATION IN CLOUD COMPUTING BASED ON GREEN CONTROL ALGORITHM

K.Ganapathy[1], M.N.Karuppusamy[2]

[1] PG Scholar, Department of Computer Science and Engineering, Sri Subramanya College of Engineering and Technology, Palani.Tamil Nadu. India.
[2] Assistant Professor, Department of Computer Science and Engineering, Sri Subramanya College of Engineering and Technology, Palani.Tamil Nadu. India.

***Abstract*--The Cloud Computing is the fast developing technology but it facing serious issues like power shortages. Here there are many critical issues for the cloud providers such as achieving an energy-efficiency control and simultaneously satisfying a performance guarantee.This implement three power-saving policies in cloud systems in order to mitigate server idle power.Cchallenges of controlling service rates and applying the N-policy to optimize operational cost within a performance guarantee. Here develop a cost function that includes the costs of power consumption, system congestion and server startups. Here demonstrate the effect of energy-efficiency controls on response times, operating modes and incurred costs. 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. Propose an algorithm called Efficient Green Control (EGC) algorithm which is developed for solving constrained optimization problems and making costs/performances tradeoffs in systems with different power-saving policies. Our simulation results show that the benefits of reducing operational costs and improving response times which can be verified by applying the power-saving policies combined with the proposed algorithm as compared to a typical system under a same performance guarantee.**

Key Words: ***Energy-efficiency control, power-saving policy, Cost optimization, response time.***

I. INTRODUCTION

As cloud computing is predicted to grow, substantial power consumption will result in not only huge operational cost but also tremendous amount of carbon dioxide (CO2) emissions. Therefore, an energy efficient control, especially in mitigating server idle power has become a critical concern in designing a modern green cloud system. Ideally, shutting down servers when they are left idle during low-load periods is one of the most direct ways to reduce power consumption. Unfortunately, some negative effects are caused under improper system controls. Burst arrivals may experience latency or be unable to access services. There has a power consumption overhead caused by awakening servers from a power-off state too frequently. The worst case is violating a service level agreement (SLA) due to the fact that shutting down servers may sacrifice quality of service (QoS) [1], [2]. The SLA is known as an agreement in which QoS is a critical part of negotiation. A penalty is given when a cloud provider violates performance guarantees in a SLA contract. An efficient green control (EGC) algorithm is first propose for solving constrained optimization problems and making costs/performances tradeoffs in systems with different power-saving policies. The challenges of controlling the service rate and applying the N-policy to minimize power consumption and simultaneously meet a response time guarantee are first studied. To address the conflict issue between performances and power-saving, a tradeoff between power consumption cost and system congestion cost is conducted.

An efficient green control (EGC) algorithm is proposed to optimize the decision-making in service rates and mode-switching within a response time guarantee by solving constrained optimization problems. As compared to a typical system without applying the EGC algorithm, more cost-saving and response time improvements can be achieved. The 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.

II. SYSTEM MODEL

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

2.1 EXISTING SYSTEM

The existing design approaches are Power-Saving in Virtual Machine, Power-Saving in Computing Infrastructure.

2.1.1 Power-Saving in Virtual Machine

In [3]. [4], Huang et al. studied the virtual machine placement problem with a goal of minimizing the total energy consumption. 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.

2.1.2 Power-Saving in Computing Infrastructure

In [5]. [6], Duggan and Young presented a basic theoretical model and used it in building managing, micro-grids, and datacenter energy management.

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 heterogeneous data centers.

Problem Identified:

- Power saving in virtual machine is not fully achieved for power consumption in cloud based storages.
- Multi User request in same time not properly response.
- More emulsion of Carbon di oxide gas due to use of server maintenance.

2.2 PROPOSED SYSTEM

Distributed service system consists of lots of physical servers, virtual machines and a job dispatcher. The job dispatcher in designed system is used to identify an arrival job request and forward it to a corresponding VM manager that can meet its specific requirements. 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.

Benefits:

- Efficient way of saving the power consumption in cloud based storages.
- Proper response to the entire requested user in the cloud computing process.
- Avoid the over emulsion of Carbon di oxide gas due to use of server maintenance.
- Practically implemented algorithm and approaches.

III. DESIGN CONSTRUCTION

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

3.1 Generation of Queuing Model

Generally Queue maintains several job request given by the authorized users. The 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). There may have some job requests that need to be performed serially at multiple service stages. Then, applying phase-type distributions allow us to consider a more general situation.

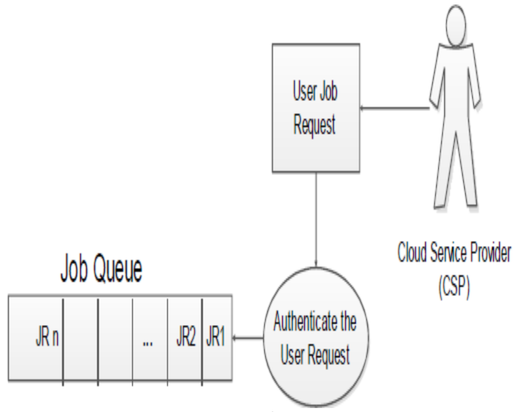

Fig-1: Generation of Queuing Model

3.2 Implementation of 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.

Two cases starting Busy Mode

Starting a busy mode when a job arrives in an idle mode; Starting a busy mode if the number of jobs in a waiting queue is more than the N value when a sleep period expires.

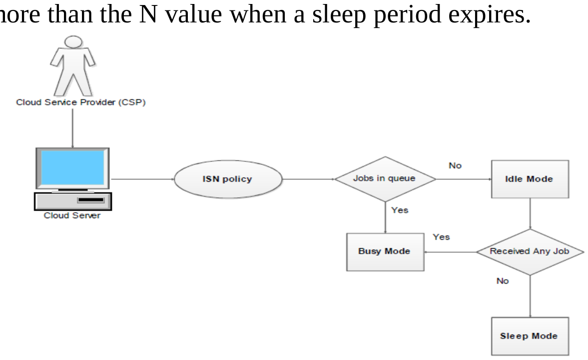

Fig-2: Implementation of ISN Policy

3.3 Modeling the 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.

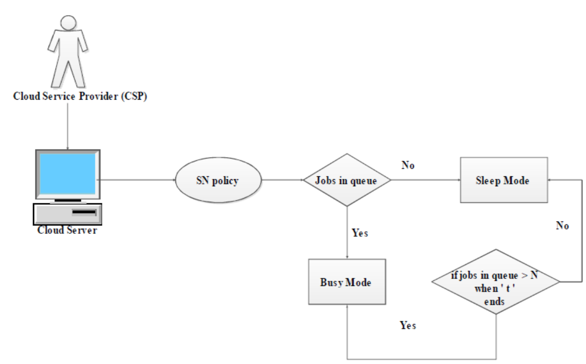

Fig-3: Modeling the SN Policy

3.4 Enhance SI Policy

According to the switching process (from Sleep to Idle), we have called such an approach “SI policy”. The step by step decision processes and job flows of the SI policy. A server switches into a sleep mode immediately instead of an idle mode when there has no job in the system. The 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. Otherwise, it switches into a busy mode without any restriction and begins to work.

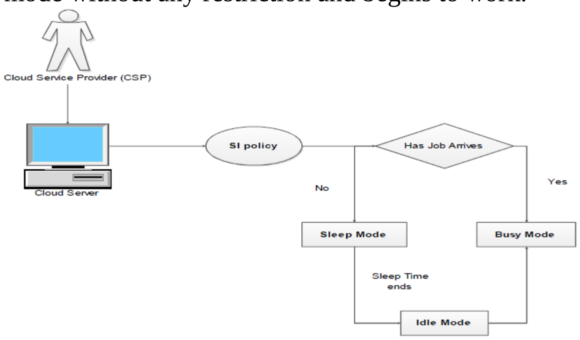

Fig-4: Enhance SI Policy

3.5 Formulation of ECG Algorithm

An EGC algorithm is presented to solve the nonlinear constrained optimization problem effectively. Meeting a SLA constraint has the highest priority, followed by cost minimization in deciding the optimal solution (M*; N*).

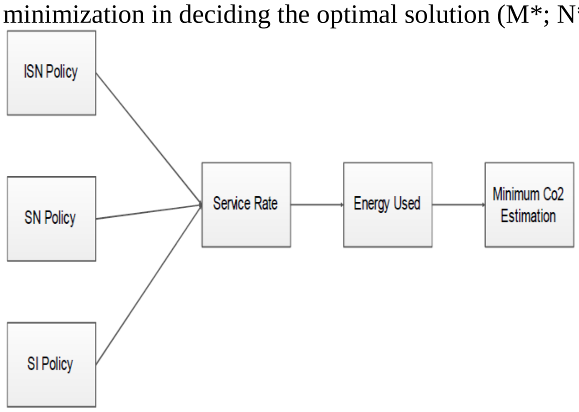

Fig-5: Formulation of ECG Algorithm

3.6 Neg late Repeating Request

The multiple users request in add the queue for cloud service provider and response for the based on any one policy. But some time repeating same user and same record in the add queue, slow down the response process and significantly increase the cost. So we can overcome this problem the cloud service provider check the repeat request in the queue using Request Matching algorithm neg late the same request for same process.

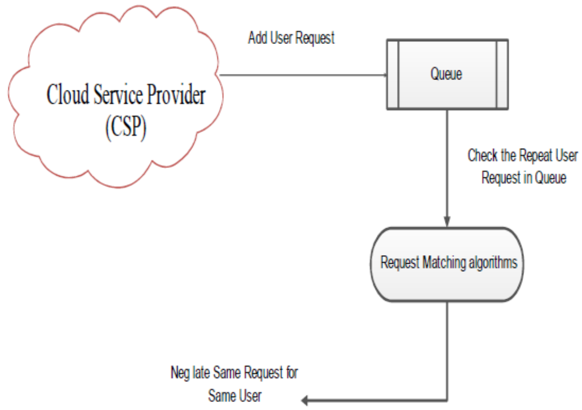

Fig-6: Neg late repeating Request

IV. PERFORMANCE EVALUATION

The system with the SI policy doesn’t reduce the sleep probability as the arrival rate increases; hence, it results in higher response times than other power-saving policies with a higher arrival rate.

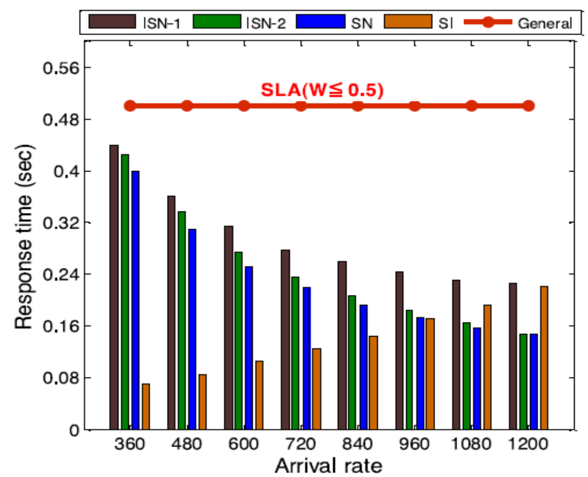

Chart-1: Response time comparison.

The Proposed power-saving policies can effectively reduce cost, especially when an arrival rate is low. For a cloud provider who focuses on reducing cost, implementing the SN policy is a better choice to deal with a wide range of arrival rates. Benefits when the startup cost is high. As compared to a general policy, cost savings and response time improvement can be verified.

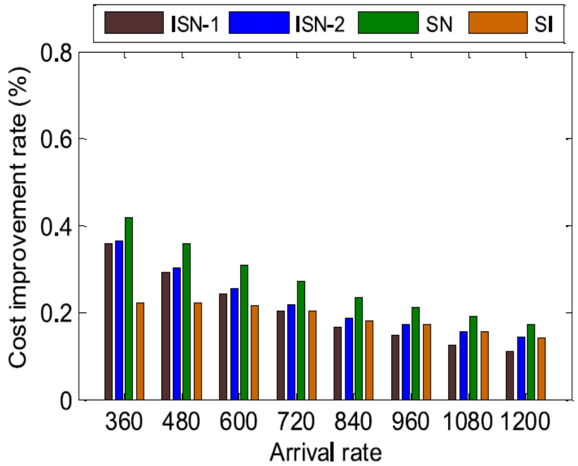

Chart-2: Operational cost improvement rates

Although the general policy tries to keep the service rate as low as possible, it still results in higher cost than other policies, as shown in finally; we measure the cost improvement rates, which calculate the relative value of improvements to the original value instead of an absolute value.

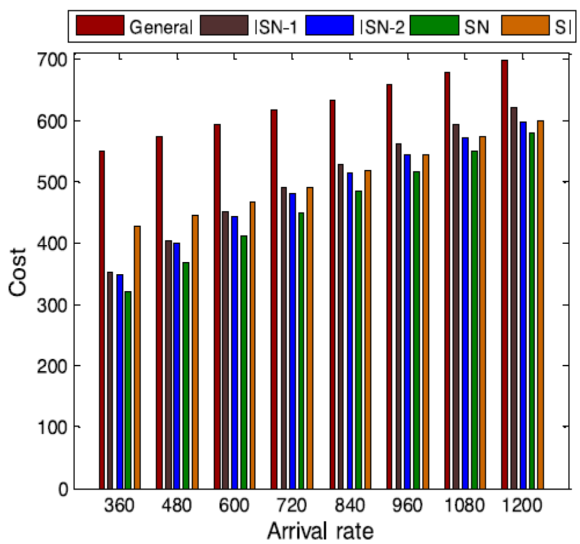

Chart-3: Operational cost comparisons.

Therefore, service rates are controlled at higher values with power-saving policies and their idle times can be red used by switching into sleep modes. Conversely, the gene real policy focuses only on a performance guarantee and reduces the service rate as low as possible for the purpose e of saving operational cost. On the other hand, since a server switches into a sleep mode only once in an operation cycle for the SI policy, the variation among sleep probabilities is slight as the arrival rate increases.

V. CONCLUSION

The Cloud Computing is the fast developing technology but it facing serious issues like power shortages. Cloud Computing is the emerging technology but the growing crisis in power shortages has brought a concern in existing and future cloud system designs. Here to mitigate unnecessary idle power consumption, three power saving policies with different decision processes and mode switching controls are considered. Our algorithm allows for cloud providers to optimize the problem in decision-making on service rate and mode-switching restriction, so as to minimize the operational cost without sacrificing a SLA constraint. The issues such as choosing a suitable policy among diverse power managements which reaches a relatively high effectiveness has been examined based on the variations of arrival rates and incurred costs. Experimental results show that a system with the SI policy can significantly improve the response time in a low arrival rate situation.

VI. FUTURE WORK

In Future work focus on the Weighted Round Robin Scheduling Algorithm allows cloud providers to optimize the problem in decision-making in service rate and mode-switching restriction, so as to minimize the Operational Cost without sacrificing a SLA constraint.

REFERENCES

[1] Amokrane , M. Zhani , R. Langar, R. Boutaba, and G. Pujolle, “Greenhead: Virtual data center embedding across distributed infrastructures,” IEEE Trans. Cloud Comput., vol. 1, no. 1, pp. 36– 49, Jan.–Jun. 2013.

[2] Y. Deng, W. J. Braun, and Y. Q. Zhao, “M/M/1 queueing system with delayed controlled vacation,” OR Trans., vol. 3, pp. 17–30, 1999. K. H. Wang and H. M. Huang, “Optimal control of an M/Ek/1 queueing system with a removable service station,” J. Oper. Res. Soc.,” vol. 46, pp. 1014–1022, 1995.

[3] G. P. Duggan and P. M. Young, “A resource allocation model for energy management systems,” in Proc. IEEE Int. Syst. Conf., 2012, pp. 1–3.

[4] 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.

[5] F. Larumbe, and B. Sanso, “A tabu search algorithm for the location of data centers and software components in green cloud computing networks,” IEEE Trans. Cloud Comput., vol. 1, no. 1, pp. 22– 35, Jan.–Jun. 2013.

[6] Y. Levy and U. Yechiali, “Utilization of idle time in an M/G/1 queueing system,” Manage. Sci., vol. 22, no. 2, pp. 202–211, 1975.