Energy Efficient Scheduling Framework for Cloud Computing using Ranking Algorithm

R.Vijindra, Lecturer, IT Department, Ranipettai Engineering College

Abstract– Cloud computing provides everything as a service and solves high performance and high throughput computing problems through sharing resources ranging from personal computers to super computers distributed around the world. In cloud computing, one of the major problems is resource scheduling, i.e. allocating tasks to resources efficiently. In addition to completion time, I also take energy consumption of data centers into account. In this paper, energy efficient scheduling framework proposed for the job scheduling problem. This framework is designed with three objectives, i.e. 1.reducing the completion time of each job, 2.reducing the energy consumption of data centers, 3.balancing the incoming load. I compare the results of the energy efficient scheduling framework with previous job scheduling algorithm implemented in the cloud environment. Simulation results show that the proposed framework perform better than existing algorithm and achieves the objectives.

Index Terms-cloud computing, job scheduling, energy efficiency, Berger model, QoS, Ranking Algorithm, Resource Management

- - - - - - - -♦- - - - - - - -

1 INTRODUCTION

In the past few years, cloud computing has emerged as an enabling technology and it has been increasingly adopted in many areas including science and engineering not to mention business due to its inherent flexibility, scalability and cost-effectiveness. A cloud is an aggregation of resources/services—possibly distributed and heterogeneous—provided and operated by an autonomous administrative body (e.g., Amazon [1], Google [2] or Microsoft [3]). Resources in a cloud are not restricted to hardware, such as processors and storage devices, but can also be software stacks and Web service instances. Recent advances in virtualization technologies and the business trend of reducing the total cost of ownership (TCO) in particular have made it much more appealing compared to when it was first introduced. Clouds are primarily driven by economics—the pay-per-use pricing model like similar to that for basic utilities, such as electricity, water and gas. While the pay-per-use pricing model is very appealing for both service providers and consumers, fluctuation in service request volume and conflicting objectives between the two parties delay its effective application. In other words, the service provider aims to accommodate/process as many requests as possible with its main objective maximizing profit; and this may conflict with consumer's performance requirements (e.g., response time) and fair resource allocation [4].

Scheduling Determines when an activity should start or end, depending on its duration, predecessor activity (or activities), predecessor relationships, resource availability, target completion time. Generic issues in scheduling which includes Fair resource allocation, Quality of service, Load balancing, Increasing Response Time, Energy consumption and so on places a vital role in cloud computing. In cloud, user may face hundreds of thousands virtualized resources for each task, so it is impossible to schedule the resources manually, Hence needed an effective framework to schedule the resources according to the user preferences along with the user satisfaction. Cloud computing is a user oriented design, where number of user growing every day, hence it should be scalable for the different number of users and user preferences. For each job, the scheduling framework should allocate the resources efficiently, so energy efficiency also places an important role in job scheduling.

There have been a number of studies exploiting market based resource allocation to tackle this problem. Noticeable scheduling mechanisms for online market include POPCORN [5], Nimrod-G [6] focuses on the market mechanisms for online resources. The first job scheduling algorithm was implemented using the

Berger model of distributed justice [7] in cloud computing area. In [7], during the job scheduling process, the algorithm establishes dual fairness constraint. The first constraint is to classify user tasks by QoS preferences, and establish the general expectation function in accordance with the classification of tasks to restrain the fairness of the resources in selection process. The second constraint is to define resource fairness justice function to judge the fairness of the resources allocation. The basic mechanism of cloud computing is to dispatch the computing tasks to resource pooling which constitutes by massive computers. It enables a variety of applications to gain computing power, storage and a variety of software services according to their needs. The commercialization and the virtualization technology adopted by cloud computing has poured into new features for cloud architecture. For example, it leaves the job scheduling complexity of cloud computing to the virtual machine layer through resource virtualization. Further, it raised a number of new features for job scheduling, such as cloud computing needs pay more attention to the fairness of resources allocation. This paper, from the fairness point of view, for the first time proposed and implemented the algorithm of job scheduling based on Berger model in cloud computing.

It is well known that high-quality service and fair competition are the foundation for commercial operations. In cloud computing, the provider needs to face a variety of tasks from different users. With the number increase of users, cloud-scale expansion and minimizing the energy, providing minimum completion time and so on, the key issues of cloud computing are to ensure the Quality of Service (QoS) and to provide, for different users, an equitable opportunity for the use of resources. The aim is to enable the user needs to get more satisfaction. In [7] the primary focus is on two QoS factors such as completion time and bandwidth. It provides distributive justice theory to provide fairness resource allocation along with the better completion time. In this paper, Berger model theory on distributive justice in the field of social distribution was first introduced into the job scheduling algorithm in cloud computing.

The remainder of the paper is organized as follows. Section 2 explores couples of various resource scheduling algorithms with respect to fair resource allocation and the energy efficiency scheduling. Section 3 describes the problem statement in detail; section 4 gives the detailed methodology and description which is used to solve the problem focused in this paper; section 5 discusses the experiment setup and the evaluation results and section 6 concludes this paper with the future directions for further research.

2 RELATED WORK

From the economic point of view, cloud computing is a new computing model which decided by economic principle. The basic philosophy of the job scheduling algorithm based on economic models is to establish market mechanisms between resource providers and resource consumers. It uses price lever to adjust user needs and resource distribution. The research on how to apply economic theory to resource distribution of distributed system can be traced back to an auction mechanism for resource allocation in the PDP-1 machine, which proposed by Sutherland in 1968 [8]. In Ref [8], an auction method is described for allocating computer time that allows the price of computer time to fluctuate with the demand and the relative priority of users to be controlled so that more important projects get better access. Auction method is based on reserving the computer time. The major limitation of this kind of methods is it has so many rules to reserve the computer time and the limited availability. Most of the subsequent research focused on solving the load balancing of distributed system with the aid of price mechanism [9]. In Ref [9], the objective was to endow processors in a distributed system with resources and behavior similar to that enjoyed by actual consumers and then allow an economic model to allocate resources between them in the manner of a normal market. The main limitations of [9] where it's mainly based on assumption about the market and the completion time and throughput parameters should be taken into the account. Along with the grid development, the researchers

on how to apply the market mechanism to grid resources allocation have also carried out in-depth [10, 11, 12, 13, 14, and 15]. Among them, Popcorn [5] and the Nimrod/G [6] are more influential. They all reveal the validity of grid resource allocation based on economics method. Ref. [16] has introduced market-oriented cloud. It is one of the first papers in the area. However, most of these studies did not involve the fairness of resources allocation under the market mechanism. Hence the Berger model of distributed justice was implemented to provide better resource allocation in cloud environment. It focuses only on fairness resource allocation; it is not focusing on the energy efficiency factor. The scheduling algorithm was implemented using the cloudsim [17] package extended version and compared with the optimal completion time algorithm. This algorithm provides better completion while comparing with the optimal completion time algorithm.

Cloud infrastructures have recently become a center of attention. They can support dynamic operational infrastructures adapted to the requirements of distributed applications. Cloud systems provide on-demand computing power and storage, so they can perfectly fit the user's requirements. But as they reach enormous sizes in terms of equipments, energy consumption becomes one of the main challenges for large-scale integration [18]. Moving to the energy efficiency parameter, various number of research works were carried out from the last decays. In ref [19], a hybrid energy-efficient scheduling approach for private clouds was proposed. In this approach, Jiandun Li et al provides an approach which is based on pre-power technique and least-load-first algorithm. This approach can save more time for users, conserve more energy and can achieve higher level of load balancing. But this approach is mainly used for daily use private clouds not for the public clouds. Jiandun Li extends his work [19] with the use of dynamic migration; proposed energy efficient scheduling algorithm for private clouds [20] to avoid the limitations of the previous work which includes 1.Intrusion is not accepted. 2. Workflow attributes regularly changes cyclically. 3. Conserving energy, whereas keeping response time too long is beyond consideration. 4. Scheduling algorithms should be light-weighted. 5. Further optimization or consolidation (e.g. migration) should be carefully applied. This paper uses simple algorithm without using any priority considerations. The use of High Performance Computing (HPC) in commercial and consumer IT applications is becoming popular. They need the ability to gain rapid and scalable access to high-end computing capabilities. Cloud computing promises to deliver such a computing infrastructure using data centers so that HPC users can access applications and data from a Cloud anywhere in the world on demand and pay based on what they use. However, the growing demand drastically increases the energy consumption of data centers, which has become a critical issue. High energy consumption not only translates to high energy cost, which will reduce the profit margin of Cloud providers, but also high carbon emissions which is not environmentally sustainable. Hence, energy-efficient solutions are required that can address the high increase in the energy consumption from the perspective of not only cloud provider but also from the environment [21]. In [21], the authors considering a number of energy efficiency factors such as energy cost, carbon emission rate, workload, and CPU power efficiency which changes across different data center depending on their location, architectural design, and management system. Clouds are essentially data centers which require high energy usage to maintain operation. Today, a typical data center with 1000 racks need 10 Megawatt of power to operate, High energy usage is undesirable since it results in high energy cost. For a data center, the energy cost is a significant component of its operating and up-front costs [21]. This paper provides various insights about the power consumption in the datacenters in cloud computing environment. Most previous work focuses on reducing energy consumption in data centers for web workloads [22][23]. Thus, they assume that energy is an increasing function of CPU frequency since web workloads have the same execution time per request. The main limitation of this paper is the overhead produced by the complex derivations and calculations. Ref [24] deals with power aware provisioning of virtual machines in real time services. The objective of the approach [24] is (i) to model a real-

time service as a real-time virtual machine request; and (ii) to provision virtual machines of datacenters using DVFS (Dynamic Voltage Frequency Scaling) schemes. According to the previous works, none of the works has been done the energy efficient scheduling framework with the fairness resource allocation.

3 PROBLEM STATEMENT

Specifically, considering previous studies and current platforms, there are at least three challenging problems in energy efficient resource scheduling in cloud computing:

How to reduce the coming request's response time,

How to balance the workloads, when the datacenter is running on low-power mode.

How to reduce the energy consumption of the data centers while providing fair resource allocation.

4 METHODOLOGY USED

In this paper, I am focusing on three main objectives as defined in the problem statement. To overcome these three issues in cloud scheduling, an energy efficient scheduling framework is proposed. According to the related work, an energy efficient scheduling framework has been proposed. The scheduling framework mainly focuses on the energy efficiency of the data centers along with the minimum completion time, fair resource allocation, and load balancing. Figure 1 shows the proposed framework architecture.

4.1 Architecture

The architecture of the proposed energy efficient scheduling framework is shown in figure.1

4.2 Module Description

4.2.1 User

User is the one who is sending requests to process and schedule. User is generating the requests based on the various criteria's like high priority jobs, low priority jobs, resource descriptions, completion time and so on.

4.2.2 Resource Broker

The resource broker components which include various components like policy prioritizer, resource match maker, database, job queue, scheduler interface, remote monitoring interface. The functions of each component are explained in the procedure.

4.2.3 Ranking Algorithm

Ranking algorithm is used for virtual machines. In a cloud environment, various numbers of virtual machines are used to access the services by using the resources. By using the ranking algorithm, virtual machines are ranked based on the resources. The functioning of the ranking algorithm is explained in the procedure.

4.2.4 Task Partitioning

Task partitioning is used to balance the loads in the virtual machines. Based on the higher priority, the jobs will be scheduled based on the resource availability.

4.3 Procedure

1. User generate the requests based upon the job characteristics and the set of performance criteria
2. The user request will send to the resource broker
3. Resource broker will perform the following functions
4. The user request will sent to the policy prioritizer to provide priority based on the policies, SLAs, if available.
5. Scheduler interface is responsible for providing priority policies, match making services of the user request, and all the users history will be saved in the database.
6. The scheduler interface contains the job queue for the low priority jobs
7. Remote monitoring interface is responsible for monitor the jobs that are in processing, available resources and so on.
8. Ranking algorithm is used to rank the virtual machines (VMs). The virtual machines are

classified under some specified categories for the fast job completion, say for example

i VMs 1-5 specifies CPU resources
ii VMs 6-10 specifies Memory resources
iii VMs 11-15 specifies Bandwidth resources
iv VMs 16-20 specifies Online resources

9. Based upon the ranking algorithm, the VMs were ranked.
10. If the user request is having priority, then the tasks will be partitioned based on the VM classifications
11. Task partitioning is used for load balancing in the cloud environment.
12. If the job doesn't have any priority, then the backfilling algorithm will be used to perform the jobs, if the VMs are totally free. If the VMs are not free, then the jobs will be sent to the job queue in the resource broker.
13. After the task partitioning stage, resource availability check will be done to ensure the resource fairness
14. If the resources available for the user requests, then the jobs will be executed based on the ranking of virtual machines.
15. If the resources are not available for the requests, then the idle VMs will be woken up to ensure the resource availability. Then the jobs will be executed.
16. If the VMs are busy with another execution, then advance reservation is used to reserve the resources for avoiding the energy consumption. Once the VMs get free, the job scheduling will be done.
17. All the results and notifications will send to the user in between to notify the user about the status of the request.
18. If no other jobs are available for processing, then the VMs are moved to idle/switch off mode to save the energy consumption.
19. VM migration is used to save the energy of data centers in this framework. Live migration greatly improves the capacities and the features of Cloud environments: it facilitates fault management, load balancing, and low-level system maintenance.

4.4 Advantages

1 Lower completion time
2 Provides high energy efficiency factor
3 High performance guaranteed by using ranking algorithm.

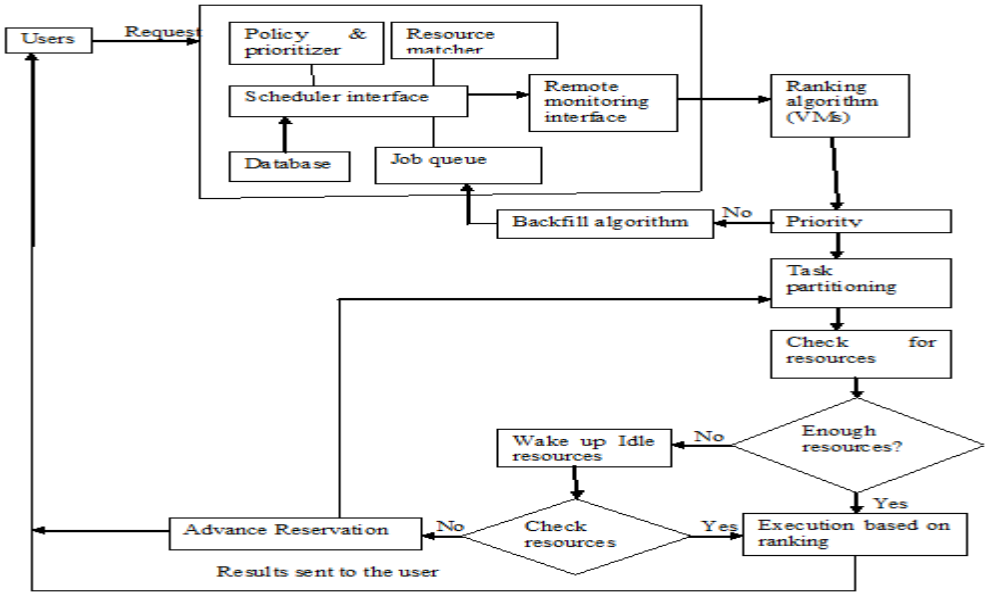

Figure 1: energy efficient scheduling framework

5 SIMULATION RESULTS

In this section, the results have been evaluated. The project is implemented using the cloudsim [17] toolkit version 2.1, the

cloudsim architecture is extended to provide the results. I firstly show simulation results for the completion time factor using Berger model of distributive justice and the energy efficient scheduling framework and the values are tabulated for the fixed amount of jobs (jobs=10). Secondly, the energy consumption of the proposed framework has shown. Figure 2 shows the completion time factor with Berger model of distributive justice. Figure 3 shows the completion time factor with the proposed energy efficient scheduling framework. By evaluating the results, the proposed framework provides better completion time compared with the Berger model of distributive justice. Figure 4 shows the energy consumption factor with respect to the proposed scheduling framework.

Figure 2: completion time factor with the Berger model of distributive justice

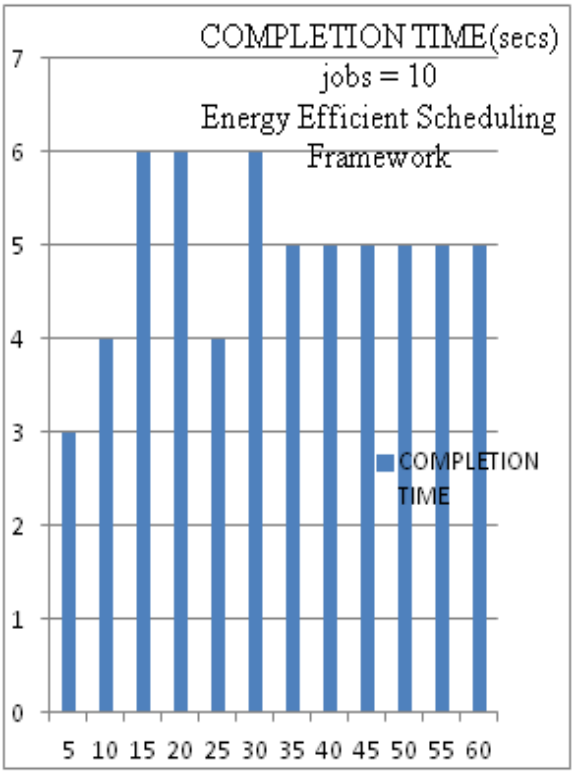

Figure 3: completion time factor with the Energy Efficient Scheduling Framework

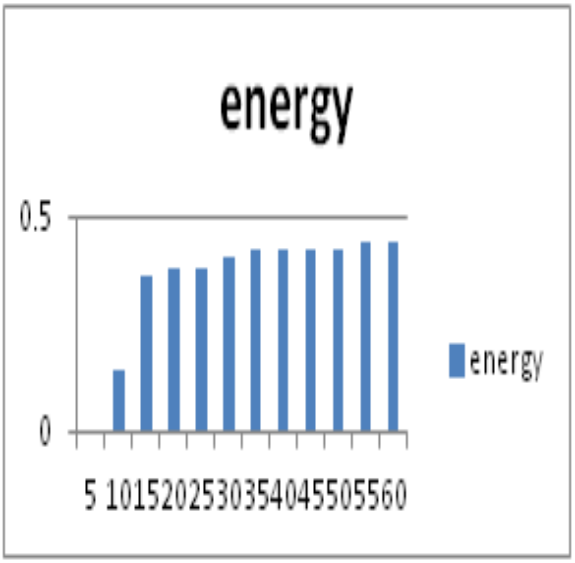

Figure 4: Energy Consumption Factor for the Proposed framework

6 CONCLUSION AND FUTURE SCOPE

Cloud Computing has emerged as one of the hot research areas in the field of computer networking. Scheduling, which decides how to

distribute tasks to resources, is one of the most important issues. Reducing energy consumption has been an essential technique for Cloud resources or datacenters, not only for operational cost, but also for improving system reliability. This paper mainly focuses on the fair resource allocation, minimizing the completion time and energy consumption while balancing the load. Ranking algorithm is used for the ranking of the virtual machines to provide faster access of resources; hence it reduces the response time. I proposed an energy efficient scheduling framework which considers the energy consumption, load balancing along with the fair resource allocation. Experimental studies shows that my proposed work shows better performance compared with the existing job scheduling algorithm based on Berger model.

The future work focuses on providing budget constraints for the resources and allocating more scheduling parameters for calculating the performance evaluation of the proposed algorithm.

REFERENCES

[1] Amazon, Aws.amazon.com
[2] GoogleApp Engine, Appengine.google.com
[3] Microsoft, www.microsoft.com
[4] Young Choon Lee, Chen Wang, Albert Y. Zomaya, Bing Bing Zhou, "Profit-driven Service Request Scheduling in Clouds", 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing,2010.
[5] Ori Regev, Noam Nisan,"The POPCORN Market – an Online Market for Computational
[6] Resources"
[7] "The Nimrod-G Grid Resource Broker and Economic Scheduling Algorithms" Chapter 4 http://messagelab.monash.edu.au/Nimrod
[8] Baomin Xu, Chunyan Zhao, Enzhao Hu, Bin Hu, "Job scheduling algorithm based on Berger model in cloud environment", journal of advances in engineering software, April 2011, 419-425
[9] Sutherland IE. "A futures market in computer time". Communication ACM publications, 1968;11(6):449–51.
[10] Ferguson D, Yemini Y, Nikolaou C. "Microeconomic algorithms for load balancing in distributed computer systems". In: Proceedings of the eighth international conference on distributed systems. San Jose: IEEE Press; 1988. p. 491–9
[11] Subramoniam K, Maheswaran M, Toulouse M. "Towards a micro-economic model for resource allocation in grid computing systems". In: Proceedings of the 2002 IEEE Canadian conference on electrical and computer engineering. Manitoba: IEEE Press; 2002. p. 782–5.
[12] Gomoluch J, Schroeder M. "Market-Based resource allocation for grid computing: a model and simulation". In: Endler M, Schmidt D, editors. Proceedings of the first international workshop on middleware for grid computing (MGC 2003). Rio de Janeiro: Springer-Verlag; 2003. p. 211–8.
[13] Braun TD, Siegel HJ, Beck N, et al. "A Comparison of eleven static heuristics for mapping a class of independent tasks onto heterogeneous distributed computing system". Journal of Parallel Distributed Computing 2001;61(6):810–37.
[14] LI Mao-Sheng, YANG Shou-Bao, FU Qian-Fei1, YANG Jin, "A Grid Resource Transaction Model Based on Compensation". Journal of Software, 2006;17(3):472–80.
[15] Broberg J, Venugopal S, Buyya R. "Market-oriented grids and utility computing: the state-of-the-art and future directions". Journal of Grid Computing, 2008;6(3): 255–76.
[16] Kumar S, Dutta K, Mookerjee V. "Maximizing business value by optimal assignment of jobs to resources in grid computing". European Journal of Operations Res 2009;194(3):856–72.
[17] Rajkumar Buyya, Chee Shin Yeo, Srikumar Venugopal. "Market-oriented cloud computing: vision, hype, and reality for delivering IT services as computing utilities". In: Proceedings of the tenth IEEE international conference on high *performance computing and* communications (HPCC 2008), Dalian, China. Los Alamitos, CA, USA: IEEE CS Press; September 25–27, 2008.
[18] Rajkumar Buyya, Rajiv Ranjan and Rodrigo N. Calheiros, "Modeling and simulation of scalable cloud computing environments and the Cloudsim Toolkit: challenges and opportunities". In: Proceedings of the seventh high performance computing and simulation conference (HPCS 2009, ISBN:978-1-4244-49071), Leipzig, Germany. New York, USA: IEEE Press; June 21–24, 2009.
[19] Anne-Cécile ORGERIE, Laurent LEFÈVRE, "When Clouds become Green: the Green Open Cloud Architecture"
[20] Jiandun Li, Junjie Peng, Zhou Lei, Wu Zhang, "An Energy-efficient Scheduling Approach Based on Private Clouds", Journal of Information & Computational Science 8: 4 (2011) 716–724.
[21] Jiandun Li, Junjie Peng, Zhou Lei, Wu Zhang, "A Scheduling Algorithm for Private Clouds", Journal of Convergence Information Technology, Volume6, Number 7, July 2011.
[22] Saurabh Kumar Garg, Chee Shin Yeo, Arun Anandasivam, Rajkumar Buyya, "Energy-Efficient Scheduling of HPC Applications in Cloud Computing Environments", ELSEVIER Publications, September 2009.
[23] L. Wang, Y. Lu, "Effcient Power Management of Heterogeneous Soft Real-Time Clusters", in: Proceedings of the 2008 Real-Time Systems Symposium, Barcelona, Spain, 2008.
[24] Y. Chen, A. Das, W. Qin, A. Sivasubramaniam, Q. Wang, N. Gautam, "Managing server energy and operational costs in hosting centers", ACM

SIGMETRICS Performance Evaluation Review 33 (1) (2005) 303{314.

[25] Kyong Hoon Kim, Anton Beloglazov, Rajkumar Buyya, "Power-aware Provisioning of Cloud Resources for Real-time Services", ACM Publications, December 2009