I J C T A, 9(9), 2016, pp. 4157-4161

Energy efficient Scheduled Load Balancing for Heterogeneous Cloud Applications

Muthulakshmi[1] and K. Somasundaram[2]

ABSTRACT

In recent days the computing industry is rapidly growing and become more popular in providing efficient interoperability in different kinds of application over different platforms. Interconnection various networks using interoperability or some other techniques to create a new environment as heterogeneous networks. Heterogeneous architectures can accommodate different core architectures with different data centers. Today's cloud computing needs an efficient design management for managing cloud data centers with improved system performance, increased energy saving and more reliability. In order to fulfill this need, this paper provides ESLB, an innovative scheduled-load balancing algorithm for allocating energy efficient resource allocation over heterogeneous systems. ESLB concentrates on the heterogeneity of the system while resource allocation task and it utilizes the entire information about the system (logical, physical and Meta information). As a result the cloud server has more resources of different types such as processing, storage, networking components and memory) are utilized in an effective manner.

Keywords: Cloud Computing, Load Balancing, Scheduling, Data Centers, Servers, Energy Efficiency.

1. INTRODUCTION

Cloud computing integrates both parallel computing and distributed computing can support computational as well as communication intensive applications. Whereas these cloud computing applications automatically provide services to the cloud users. Services can be provided in terms of software (SAAS), platform (PAAS) and hardware (IAAS). Services provided from small scale to large scale applications. But the most common problem occurs in cloud computing is resource allocation which is not in efficient manner. Efficiency says time, accurate resource against the user request. One of the best way to provide service oriented business to the entire world can be achieved by cloud computing. The services are in the form of providing data accessibility, various computations on data, and large size storage services with low cost. Flexibility, reliability, accuracy, fast and low cost are the benefits of cloud services. Most of the upper middle size companies are using cloud clusters in pay-and-use method [1]. The data centers are implemented in a geographically distributed manner and these are energy hungry, consuming 2% of world electricity [2]. Many research people discussed about the energy consumption, whereas in data centers only 40% of the energy is consumed by IT equipment [3]. The remaining energy is shared by the resources needed for cooling the processor (45%) and dispatching the power (15%). Same time computer servers consumed 70% and all the equipment required for communication consumed (30%) [4]. Due to the fast growth of the IT industry designing and developing energy efficient techniques are most essential for data centers are increasing proportionally [5, 6]. Similar to energy concern scheduling is also one of the parameters can improve the QoS in cloud computing.

Dzmitry Kliazovich et al. (2015) [7] proposed a novel communication model as CA-DAG, developed in direct acyclic graph. CA-DAG model represents set of computing devices is vertices and edges are

[1] Assistant professor, Department of Computer Science and Engineering, A.V.C. College of Engineering – Mayiladuthurai, Tamil Nadu, India, *Email: kavithakamalan240101@gmail.com*

[2] Professor/CSE, Vel Tech University, Chennai, Tamil nadu, India, *Email: soms72@yahoo.com*

communications. This model gives more ability to provide scheduling for resource allocation in terms of optimized priority. Various earlier research works were discussed about scheduling like delay model [7], classical scheduling, communication delays while parallel processing tasks in a same processor [7] and large communication delays [8-11]. Even through various approaches are already existing for scheduling the accuracy of scheduling is not yet fulfilled up to the market for high load.

One of the main limitation presented in the existing system called as HEROS [12] is the decision making process is less complex. In the existing system the order of complexity is defined as N, where N is the size of the job. From the background study it is clear that, one of the most essential need in cloud computing is to provide a better service in terms of resource allocation. The efficiency is measured by the time taken for resource allocation and the accuracy of the resources matched to the user request. This paper is motivated to provide a better solution for cloud computing through efficient scheduled load balancing for cloud computing. The main contribution of this paper is:

- Load balancing
- Scheduling
- Reduce the energy consumption
- Allocate the resources

2. ESLB: EFFICIENT SCHEDULED LOAD BALANCING

It is considered that, the number of incoming jobs is *N*. It is assumed that the job scheduling (JS) problem is considered in distributed infrastructures.

The incoming jobs are directly sent to the server where the SLB handles the allocation process physically or virtually. Various kinds of resources available in the cloud and they are clustered according to the resource type. All the resources can be verified by their unique characteristics and some quantitative values. The proposed server architecture is depicted in Fig. 1.

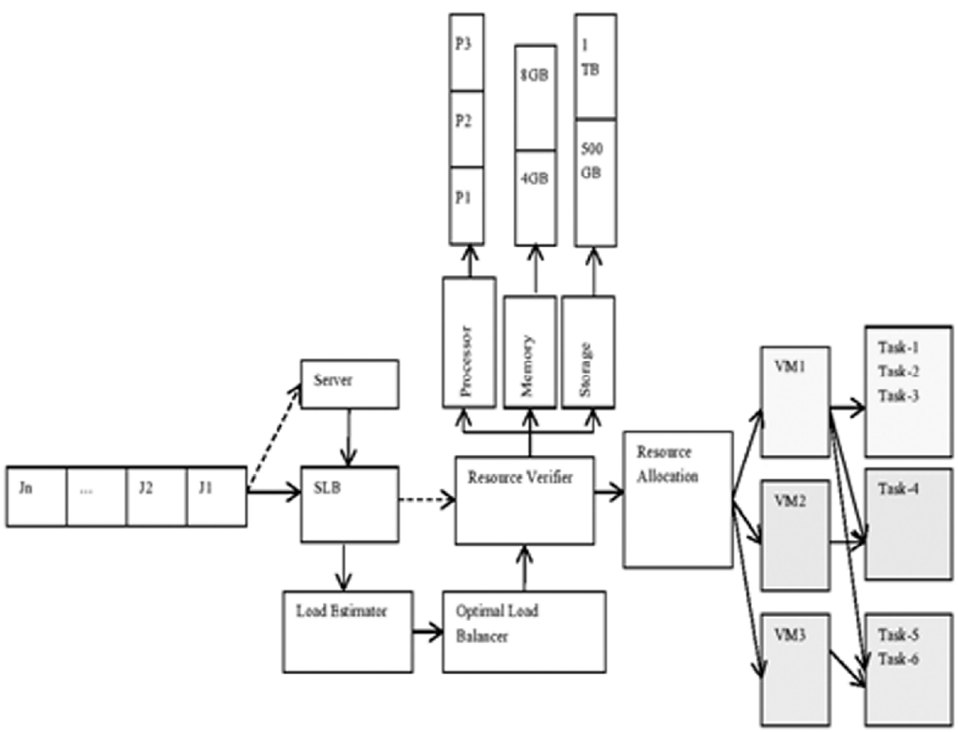

Figure 1: Resource Allocation for Heterogeneous resource

Whenever a new job comes in a queue, the SLB allocates the server according the characteristics and capabilities of the resources. This information is pre-stored in the server. Each job is an individual process which is allocated to the resources according to their description and input parameters. According to the data volume the resources are allocated by their capabilities. This kind of resource allocation and utilization provides more efficiency in resource allocation and utilization. The incoming jobs are arranged in a normal job-queue. After scheduling the queue is termed as scheduled queue. It is well known that cloud environment has interconnected by a number of servers like google.in, google.my etc. During the resource allocation process, SLB chooses a server which is connected to the data center with the highest available bandwidth. These servers can be allocated to accommodate the scheduled jobs in scheduled queue.

Sometime one job may require two or more resources like memory and storage. In these scenarios the energy efficient scheduling in required for cloud computing. In order to achieve minimum energy consumption the idle servers are assigned into a sleep state. Whenever a jobmatches, the server can be changed into an active state. Also the power management is done by using low-power components at the maximum number of times. Similarly the response time is determined as the time difference between the input of job and output of a job. If N numbers of jobs (J_i) are assigned in a queue, SLB takes the jobs on by one and group them according to the requirement. G_i represents the group-i which has unique characteristics and they can be assigned to same servers by virtualization. Scheduling can be assigned by the load of the job. If the job is mass in volume, then it should be computed first by an efficient server. Efficient server can be obtained by investigating and selecting by highest available bandwidth server connected to the data center. The performance power consumption (PPc) can be calculated by computing the performance of the job (perf(J)) and the power consumption of the load Pc(l).

$$PPc(l) = per\,f(Ji)/Pc(load) \tag{1}$$

Also the server can be selected using the following equation as:

$$ESLB(J) = PPc(JL).\left(1-\gamma.\frac{1}{1+e-\frac{\alpha}{\max\, jl_s}\left(l-beta.\max\, jl_s\right)}\right) \tag{2}$$

ESLB(l) denotes the energy efficient scheduled load balancing for job J, max(jl) is the maximum server load assigned for job J and PPc(JL) performance power consumption for job jl. All these scheduling, energy consumption and load decides the complete efficacy of the cloud computing. The main complex decision making task is to select the server to complete the job. In case of not able to decide, any, available, closer to data center and not job assigned server can be selected randomly. Similarly the efficiency can be improved in terms of load balancing where the heavy load is assigned in cloud racks.

3. EXPERIMENTAL RESULTS AND DISCUSSION

One of the modern cloud computing simulators called is GreenCloud Simulator, where it provides a cloud environment on top of Network Simulator. Various functionalities and energy efficiency of cloud data centers and its communication can be modeled and investigated in GC. Energy consumption can be obtained by simulating various elements like servers, nodes, switches, routers, selecting routing protocols and possible network links. GC simulator supports a three tier architecture based data center simulation like BCube, FiConn, DPillar and so on. The interconnecting heterogeneous server is also possible in GC. In this paper GreenCloud simulator is used to carry out the entire simulation of ESLB. ESLB algorithm is simulated in GreenCloud environment and it is validated by comparing the obtained results with the existing approaches such as HEROS and DENS. HEROS and DENS are a set of benchmarks. The GC simulator is integrated with inbuilt schedulers such as Round Robin, Random and Green scheduler where these schedules are the reference schedulers coded in GC simulator. Round Robin and Random schedulers make uniform decision and the cyclic method for allocating a processor. The other two schedulers use greedy selection method, since implementing ESLB in GC in more efficient. In order to simulate the ESLB procedure, there are two different scenarios are considered and different parameters are assigned in GC for simulation. The scenarios taken for simulating ESLB is also benchmarked scenarios. There are two different scenarios defined here are small size scenarios and large size scenarios. The configuration is experimented in 50 individual rounds of runs. Figure-2 to Figure-5 represents the mean values denoting the qualities and the corresponding relative values. The main motto of comparing the scheduler is computing the server energy consumption and mean response time. ESLB obtained a low energy consumption and low response time. The relative value obtained by ESLB, HEROS and DENS are graphically illustrated in Figure-2 to Figure-5. ESLB obtained less mean response time and less energy consumption than HEROS and DENS for all the four scenarios such as small size homogeneous, small size heterogeneous, large size homogeneous and large size heterogeneous.

The small variation among Fig-2 to Fig-5 is, due to homogeneity the energy consumption and the response time taken by all the algorithms are less comparing with the heterogeneity. Because of heterogeneous the scheduler cannot verify the feasibility of their resource allocations. Hence, it is proved that resource allocation is little complex for heterogeneous cloud computing resources. But still ESLB proved that it is efficient for reducing the energy consumption, and feasible scheduling and load balancing in cloud computing environment.

4. CONCLUSION

The ESLB algorithm is a combined work taken from the state of the art works provided solution in terms of energy efficiency, scheduling and load balancing. ESLB is specially designed to provide a better solution

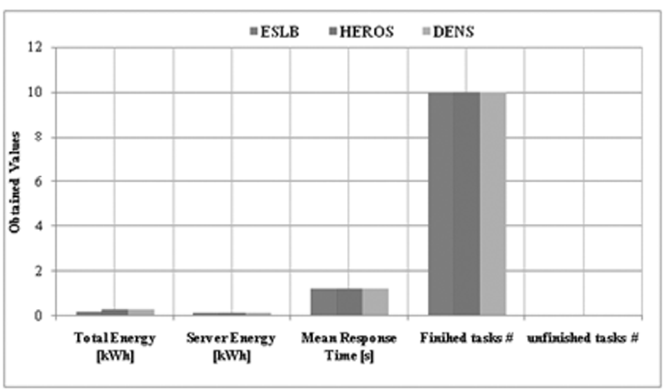

Figure 2: Small homogeneous topology – relative results

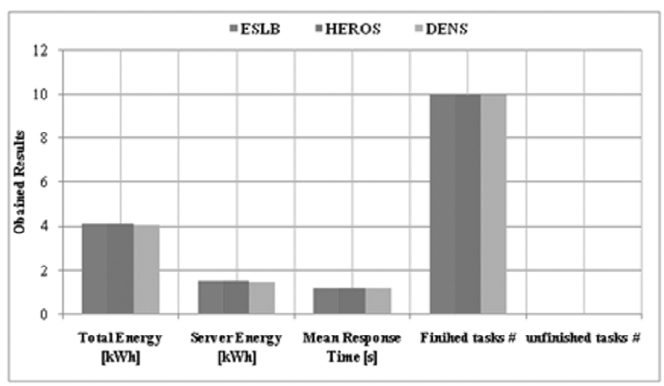

Figure 3: Full-Scale Homogeneous Topology – Results

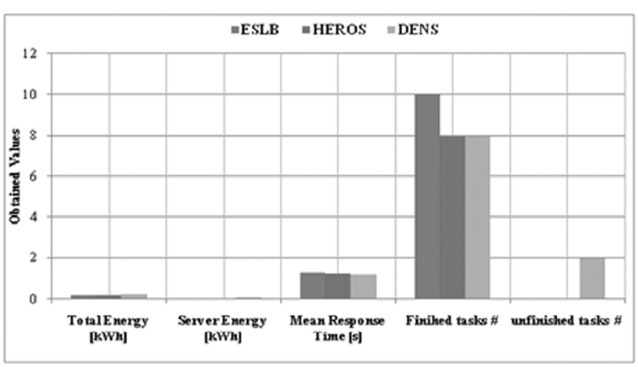

Figure 4: Small heterogeneous topology – relative results

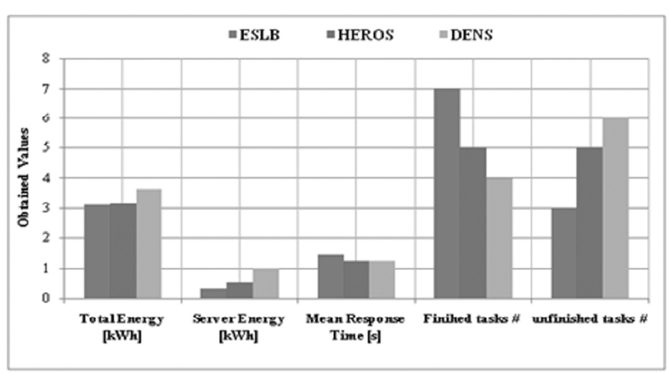

Figure 5: Full-scale heterogeneous topology – relative results

for cloud computing over heterogeneous computing. The decision making in terms of load, priority, processor characteristics and virtualization of the node are automatically carried out in ESLB. The power consumption is reduced by low computing resource utilization, scheduling by highest load and available resources and less time for completing the job requirement. The results obtained effectively from ESLB are shown in Fig-2, Fig-3, Fig-4 and Fig-5 for small and large size homogeneous, heterogeneous networks respectively. From the results it is proved and concluded that ESLB can provide energy efficient scheduling with load balancing for cloud computing environments.

REFERENCES

[1] M. Rosoff, “Here’s Why Cloud Computing Is So Hot Right Now”, 2011.

[2] “Report to congress on server and data center energy efficiency: Public law 109-431,” Lawrence Berkeley National Laboratory, 2008.

[3] Li Shang, Li-ShiuanPeh, and Niraj K. Jha, “Dynamic Voltage Scaling with Links for Power Optimization of Interconnection Networks,” Proceedings of the 9th International Symposium on High-Performance Computer Architecture table of contents, 2003.

[4] J. Liu, F. Zhao, X. Liu, and W. He, “Challenges Towards Elastic Power Management in Internet Data Centers”, WCPS 2009, in conjunction with ICDCS 2009., Montreal, Quebec, Canada, June 2009.

[5] L. Wang and S. U. Khan, “Review of Performance Metrics for Green Data Centers: A Taxonomy Study,” Journal of Supercomputing, 2011.

[6] J. Kolodziej and S. U. Khan, “Data Scheduling in Data Grids and Data Centers: A Short Taxonomy of Problems and Intelligent Resolution Techniques,” LNCS Transactions on Computational Collective Intelligence, 2011.

[7] DzmitryKliazovich · Johnatan E. Pecero et al., “CA-DAG: Modeling Communication-Aware Applications for Scheduling in Cloud Computing”, Springer, J Grid Computing, DOI 10.1007/s10723-015-9337-8.

[8] Kliazovich, D., Bouvry, P., Khan, S.U, “DENS: Data Center Energy-Efficient Network-Aware Scheduling”, Cluster Computing, 16(1), 65–75 (2013).

[9] Lep‘ere, R., Trystram, D.: A new clustering algorithm for large communication delays. IPDPS (2002).

[10] Kliazovich, D., Arzo, S.T., Granelli, F., Bouvry, P., Khan, S.U.: e-STAB: Energy-Efficient Scheduling for Cloud Computing Applications with Traffic Load Balancing. In: IEEE International Conference on Green Computing and Communications (GreenCom), Beijing, China, pp. 7–13 (2013).

[11] Guzek, M., Kliazovich, D., Bouvry, P.: A Holistic Model for Resource Representation in Virtualized Cloud Computing Data Centers. In: IEEE International Conference on Cloud Computing Technology and Science (CloudCom), Bristol, UK (2013).

[12] Guzek, Mateusz, Dzmitry Kliazovich, and Pascal Bouvry. “HEROS: Energy-efficient load balancing for heterogeneous data centers.” *2015 IEEE 8th International Conference on Cloud Computing*. IEEE, 2015.