Energy-Saving Resource Allocation in Cloud Data Centers

Moh Moh Than[1]
University of Computer Studies, Yangon[1]
mohmohthan@ucsy.edu.mm[1]

Thandar Thein[2]
University of Computer Studies (Maubin)[2]
thandartheinn@gmail.com[2]

Abstract

As the demands of cloud computing intensify, more facilities running power-intensive servers are necessary for cloud data centers. Such data center servers consume a large amount of energy, resulting in high operating costs. Therefore, high energy consumption has become a growing problem for cloud data centers. This paper proposes two energy-saving resource allocation algorithms that take into account several energy efficiency factors: resource allocation policies, power management techniques, and power models for better energy management in cloud data centers. These two algorithms are implemented on CloudSim toolkit and evaluated with real-world workload data. By the experimental evaluation of these algorithms for competitive analysis of energy consumption, we found that this work contributes to save the energy consumption in cloud data centers.

Keywords: *Cloud data centers, CloudSim, Energy consumption, Power management technique, Resource allocation*

1. Introduction

Cloud computing has become an essential aspect of the modern IT world, and cloud data centers become growing exponentially. Data centers' power consumption has increased in recent years due to an increase in size and number of data centers. The datacenter that delivers cloud services contains tens of thousands of servers. Data centers take advantage of virtualization technology [1] to host multiple virtual machines (VMs) on a single physical server. Electricity cost for powering servers forms a significant portion of the operational cost of data centers.

It is expected that the electricity demand for data centers to rise more than 66% over the period 2011–2035 [2]. It is estimated that energy costs may contribute even more than the cost of IT in the near future. Cloud service providers need to implement energy efficient management of data center resources to meet the increasing demand for cloud computing services and ensure low costs. Hence, there is a growing interest in saving energy consumption of cloud data centers.

Due to massive power consumption levels of data centers, energy-saving techniques have become essential to maintain both energy and cost-efficiency. Resource allocation is the most critical tasks in cloud computing. It involves identifying and allocating resources to every user request in such a manner that user requirements are met, and the goals of the cloud service providers are satisfied. These goals are primarily related to energy consumption or cost-saving.

Power management is also important since effective power management improves energy efficiency. As servers are the primary consumers of power in a data center, power management techniques are used to minimize power consumption by shutting temporarily down servers when they are not utilized. This research work also focuses on four energy-aware power models on cloud infrastructure environment.

CloudSim [3] is a generalized and extensible simulation toolkit that facilitates cloud infrastructures to be experimented and modeled. It supports for user-defined policies to allocate virtual machines to hosts. It is written in Java programming language and supports built-in classes to simulate the cloud environment on a single computing node. As a cloud environment consists of a large number of nodes, creating a real cloud and testing our proposed algorithms on it is impossible. So CloudSim offers a very suitable simulation environment to experiment our proposed energy-saving resource allocation algorithms.

This paper aims to manage data center power consumption and energy usage. In this paper, three different power management techniques and four energy-aware power models for two allocation policies are compared and analyzed to choose the most energy-efficient one. Based on these energy consumption comparisons, two energy-saving resource allocation algorithms are proposed and compared to gain higher energy efficiency. Cloud computing environments are simulated on the Cloudsim toolkit.

The remaining part of the paper is arranged as follows. Section 2 presents the literature review, section 3 discusses heuristics for energy efficient management, section 4 presents simulation setup, section 5 describes analysis for energy consumption of two allocation policies, section 6 presents proposed energy-saving resource allocation algorithms, and finally section 7 provides conclusion.

2. Literature Review

This section discusses state-of-the-art researches and technologies related to energy-saving resource allocation that eliminates a large portion of energy consumption in cloud data centers.

Ali et al. [4] proposed Energy Efficient VM allocation algorithm for data centers by selecting the most energy efficient host first. To reduce the power consumption in data centers, they applied three power management techniques: non power aware (NPA), power aware (PA) and dynamic voltage frequency scaling (DVFS) to their algorithms. Their algorithm achieved 23%, 23% and 9% more power efficiency than Best Resource selection (BRS), Round Robin (RR) and Minimum Power Difference (MPD) algorithms.

In [5], the researchers enhanced Round Robin (RR) algorithm by reintegrating Round Robin with shortest job first (SJF) algorithm that are selected for processing according to shortest task firstly in RR fashion then select the optimal job. This Scheduling algorithm gives a better result compared to RR.

Beloglazov et al. [6] proposed an energy-aware VM allocation algorithm that provision and schedule cloud data center resources to the user’s tasks in an efficient manner. They demonstrated that their proposed algorithm reduces the level of data center’s energy consumption.

An interior search based VM allocation algorithm: Energy-Efficient Interior Search (EE-IS) is proposed by the authors [7] for saving energy consumption and proper resource utilization. The proposed algorithm is implemented and tested on CloudSim and compared the amount of energy consumption with Genetic Algorithm (GA) and Best-fit Decreasing (BFD) algorithm. They showed that average 30% of energy has been saved using their proposed EE-IS as compare to the energy consumption of GA and BFD.

The research paper [8] evaluated the performance of four different power models: square root, linear, square and cubic models over IaaS Cloud infrastructure. They verified that cubic power model consumes less power than other three models.

This paper proposes two energy-saving resource allocation algorithms: DVFS enabled first come first serve (DFCFS) and DVFS enabled shortest job first (DSJF) considering energy efficiency factors: allocation policies, power management techniques and power models. We extend CloudSim to enable energy-saving resource allocation for data centers and to evaluate the performances of the proposed algorithms.

3. Heuristics for Energy Efficient Management

The total energy consumption is calculated by multiplying the power and time needed to turn on the servers. The energy efficiency objective aims to minimize the power consumption of servers in cloud data centers. Several energy-saving techniques can be used to monitor and control energy consumption.

3.1 Resource Allocation Policies

Resource allocation is the process of creating VM instances that match with the incoming requests onto hosts (servers). This paper emphasizes the following two allocation policies.

First Come First Serve (FCFS) – It is the simple allocation strategy. The request which comes first to the data center is allocated to the VM first. The only data required by allocator to make allocating decision is the arrival time of the request.

Shortest Job First (SJF) - The request is allocated to the VM with least run time among the requests in the ready queue. The request is always assigned to the process with least run time requirement. If the two requests having the same length, next request to be allocated, FCFS scheduling is used i.e. one which arrives first, will be allocated first to VM.

3.2 Power Management Techniques

Several techniques have been proposed for managing power consumption of data centers. The following are the existing techniques used for reducing power consumption without degrading the performance of servers in data centers:

Non Power Aware (NPA) - It calculates the server's energy usage without using any energy-saving method. Since there is no energy-saving mechanism, the total energy is dependent on the power consumption of the switched on servers and is independent of the CPU utilization. The servers use the same amount of maximum power for both levels of extremely low and high CPU usage.

Power Aware (PA) - It calculates the energy consumption of servers independent of the CPU utilization as NPA and the utilized servers consume the same amount of maximum power for both extremes of too low and too high CPU utilization. It supports to shut down unused machines in the data center.

Dynamic Voltage and Frequency Scaling (DVFS) - It is able to save power consumption of a CMOS integrated circuit, a modern computer processor. The most of power consumption in CMOS circuits is composed of static and dynamic power. The dynamic power consumption is defined by multiplying the voltage square with system frequency as in (1). It scales the power of the system varying both CPU frequency and voltage. System voltage and frequency of a server can be adjusted by DVFS technology [9] without restart the power.

$$P_d = a * c * v^2 * f \qquad (1)$$

where P_d is dynamic power consumption, a is switching activity, c is capacitance, v is voltage, and f is frequency.

Fan et al. [10] have found a strong relation between the total power consumption of a server and CPU utilization that power consumption of a server grows linearly with the growth of CPU utilization. DVFS is the dynamic power management technique [11] which reduces the dynamic power consumed by dynamically changing the frequency and the voltage of the processor during execution depending on the CPU utilization. By the time CPU voltage is decreased depending on the CPU utilization, a huge amount of energy is saved. Therefore, the cloud service providers can increase their profit by reducing the dynamic power consumed.

3.3 Power Models

By using the utilization of the CPU server and its power consumption in idle and maximum states, the power consumption of CPU servers can be estimated by power models. CloudSim provides an abstract "Power Model" implementation that can be extended to support various power models [12]. In recent releases of CloudSim, the provided power models are as follow:

Linear model: $P(u) = P_{idle} + (P_{max} - P_{idle}) * u \qquad (2)$

Square model: $P(u) = P_{idle} + (P_{max} - P_{idle}) * u^2 \qquad (3)$

Cubic model: $P(u) = P_{idle} + (P_{max} - P_{idle}) * u^3 \qquad (4)$

Square root model: $P(u) = P_{idle} + (P_{max} - P_{idle}) * \sqrt{u} \qquad (5)$

where current CPU utilization u, the maximum power value P_{max} and the idle power value P_{idle} of the CPU server.

4. Simulation Setup

CloudSim toolkit [3] is used to simulate the virtualized cloud computing environment. We extend CloudSim to allow energy-saving resource allocation for cloud data centers.

The real-world workload traces chosen for the experimentation is the RICC dataset (RIKEN Integrated Cluster of Clusters), publicly available at Parallel Workload Archive [13]. There are enormous collections of workload traces (dataset) of a variety of High Performance Computing. It contains trace of several thousands of submitted job requests over a period of five months; each has arrival time, run time (length), amount of requested CPU and memory.

To evaluate the proposed algorithms, we consider a cloud infrastructure provider with 3 data centers and, each data center has 45 heterogeneous physical servers with five different configurations shown in Table 1.

Table I. Server types characteristics

Server Type	Number of Core	Power	Number of Servers
Type 1	64	1100 W	9
Type 2	64	750 W	9
Type 3	32	750 W	9
Type 4	28	800 W	9
Type 5	16	750 W	9

In this work, we modify CloudSim to enable the simulation of High Performance Computing jobs from the logs of workloads on parallel machines [13]. At the start of simulation, VMs are created, and jobs (cloudlets) are submitted to data center broker which maps a job to a VM. Since we focus on allocation of VMs to physical hosts, we create a one-to-one mapping between cloudlets and VMs. Moreover, we implement VM termination during simulation to ensure complete simulation. The allocation of VMs to hosts utilizes our proposed algorithms. Space shared policy is used to assign the cloudlets to VMs, so that the jobs are sequentially executed in each VM. Using this policy, each job unit has its own dedicated core, and thus number of incoming jobs or queue size did not affect the execution time of individual job units since the proposed algorithms use non-preemptive method. Power management techniques and power models are implemented using the class PowerModel. This class offers getPower() function, that, returns the power consumption of hosts depending on power management techniques and power models. CloudSim's inner code is modified to evaluate our proposed algorithms and to compare them. Then our own allocation classes are specified to extend the basic CloudSim classes.

5. Analysis for Energy Consumption of Two Allocation Policies

Three power management techniques: NPA, PA and DVFS, and four energy-aware power models: square root, linear, square and cubic are applied in two allocation policies: FCFS and SJF. Energy consumptions are compared and analyzed in each allocation policy to choose the most efficient one for better energy management in cloud data centers.

5.1 Energy Consumption Comparison in FCFS Resource Allocation

The comparison of energy consumption with NPA, PA and DVFS techniques in FCFS allocation policy for different number of requests are shown in Figure 1. DVFS has lower power consumption in comparison with PA and NPA because these two mechanisms use maximum power for all servers although PA supports to shut down the servers which are not used and NPA does not support to shut down the unused servers. DVFS consumes power depending on the CPU utilization of the utilized servers.

Table 2. provides energy consumption (kWh) with various power models of DVFS in FCFS allocation policy for the different number of requests. It can be seen that the cubic power model is the most effective one compared to the other three models for every number of requests. This model can, therefore, be used to save energy in data centers.

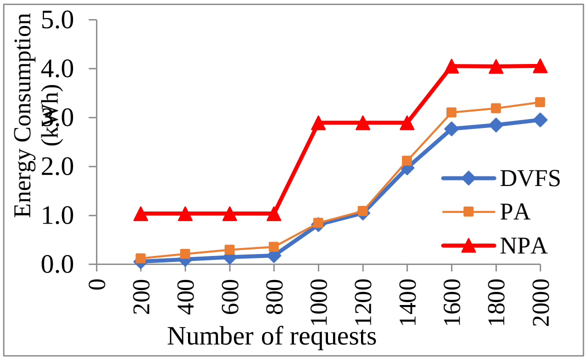

Figure 1. Comparison of energy consumption with different power management techniques in FCFS resource allocation policy

Table II. Energy consumption comparison of DVFS with four power models in FCFS resource allocation policy

Number of requests	**Square root**	**Linear**	**Square**	**Cubic**
200	0.05716	0.05716	0.05666	**0.05659**
400	0.10476	0.10361	0.10275	**0.10254**
600	0.14685	0.14611	0.14538	**0.14510**
800	0.17843	0.17778	0.17700	**0.17661**
1000	0.81771	0.81339	0.80895	**0.80704**
1200	1.05271	1.04889	1.04533	**1.04371**
1400	1.97530	1.97027	1.96196	**1.95536**
1600	2.77189	2.77059	2.77025	**2.77023**
1800	2.84846	2.84661	2.84552	**2.84533**
2000	2.95514	2.95378	2.95194	**2.95084**

5.2 Energy Consumption Comparison in SJF Resource Allocation

The comparison of energy consumption with NPA, PA and DVFS techniques in SJF allocation policy for different number of requests are presented in Figure 2. It can be clearly seen that DVFS consumes less power than other two techniques NPA and PA because these two techniques use maximum power for all servers when the servers are on while DVFS consumes power depending on the CPU utilization, and shuts down the servers that are not used. Thus, a huge amount of energy is saved with DVFS.

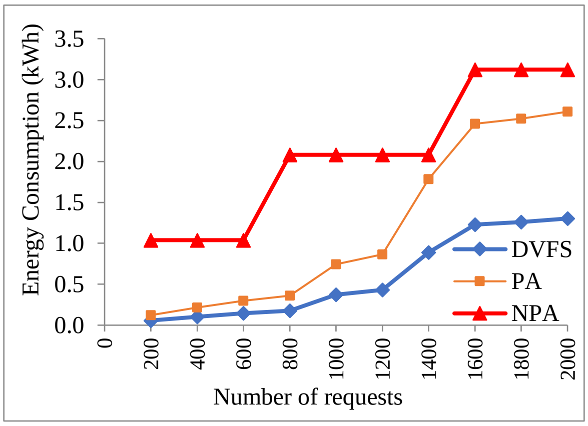

Figure 2. Energy consumption comparison of different power management techniques in SJF resource allocation policy

Table 3. presents energy consumption (kWh) with different power models of DVFS in SJF allocation policy for different number of requests. It can be found that energy consumption with the cubic power model is less than the other three models. Hence this model is applied in the proposed algorithms for better energy efficiency in the cloud data centers.

Table III. Energy consumption comparison of DVFS with four power models in SJF resource allocation policy

Number of requests	Square root	Linear	Square	Cubic
200	0.05821	0.05716	0.05666	**0.05659**
400	0.10476	0.10361	0.10275	**0.10254**
600	0.14685	0.14611	0.14538	**0.14510**
800	0.17843	0.17778	0.17700	**0.17661**
1000	0.37145	0.37133	0.37110	**0.37090**
1200	0.43174	0.43095	0.42990	**0.42927**
1400	0.89015	0.88819	0.88470	**0.88168**
1600	1.22865	1.22711	1.22519	**1.22418**
1800	1.26036	1.25903	1.25724	**1.25616**
2000	1.30356	1.30326	1.30276	**1.30235**

6. Proposed Energy-Saving Resource Allocation Algorithms

In this paper, two energy-saving resource allocation algorithms: DFCFS and DSJF algorithms are proposed based on the results of the above experiments. The energy consumptions of these two algorithms are compared to choose the better energy efficient one.

6. 1 Proposed DFCFS Algorithm

Energy-saving resource allocation DFCFS algorithm shown in Figurer 3. is proposed based on FCFS scheduling tasks and DVFS power management technique with cubic power model.

```
Algorithm1: DFCFS Algorithm
procedure Resource Allocation ($VM_j$, $Host_i$)
Add user request to VM based on matching configuration
for all VM from j to n do
    for all Host from i to m do
        if $VM_j$ fits in $Host_i$ then
            Calculate the remaining CPU capacity of Host after VM has been added
        end if
    end for
    Start a new Host in the data center, and allocate remaining VM into a new Host
end for
Calculate total energy consumption of servers with DVFS and cubic power model
end procedure
```

Figure 3. DFCFS Algorithm

6.2 Proposed DSJF Algorithm

In DSJF algorithm shown in Figure 4, the resource allocation policy SJF is applied and the energy consumption of the servers are calculated by using DVFS with cubic power model chosen according to the above comparative and analysis results.

```
Algorithm2: DSJF Algorithm
procedure Resource Allocation ($VM_j$, $Host_i$)
Sort the request list in ascending order according to their length
for all requests in the sorted List do
    Add user request to VM based on matching configuration
    if the current request length = next request length then
        Sort the request according to their arrival time
    end if
end for
for all VM from j to n do
    for all Host from i to m do
        if $VM_j$ fits in $Host_i$ then
            Calculate the remaining CPU capacity of Host after VM has been added
        endif
    end for
    Start a new Host in the data center, and allocate remaining VM into a new Host
end for
Calculate total energy consumption of servers with DVFS and cubic power model
end procedure
```

Figure 4. DSJF Algorithm

6.3 Energy Consumption Comparison of DFCFS and DSJF

Figure 5. shows comparison of energy consumption with DFCFS and DSJF algorithms for different number of requests. The energy consumptions for DFCFS and DSJF from 200 to 800 numbers of requests are almost indistinguishable. The reason behind is that the run time of those requests are similar in spite of sorting ascending order as DSJF algorithm. As the algorithms mentioned above, DSJF sorts ascending order for the run time (length) of requests before execution, while DFCFS does not. There is no impact of DSJF algorithm for the requests having almost equal run time. When the requests over 800 having different run time, it can be seen the impact

of DSJF algorithm that can save energy consumption compared to DFCFS algorithm. DSJF takes minimum turnaround time because the shortest length task gets finish in shortest possible time. Then the VM that has now become idle after executing the current task can take up next selected task. This will minimize the number of active VMs as well as active servers so that power consumption is less. In this way, DSJF resource allocation can greatly reduce power consumption. DSJF algorithm gains higher energy efficiency for the case where run time of incoming requests are different from each other and it can save up to 55% of energy consumption compared to DFCFS algorithm.

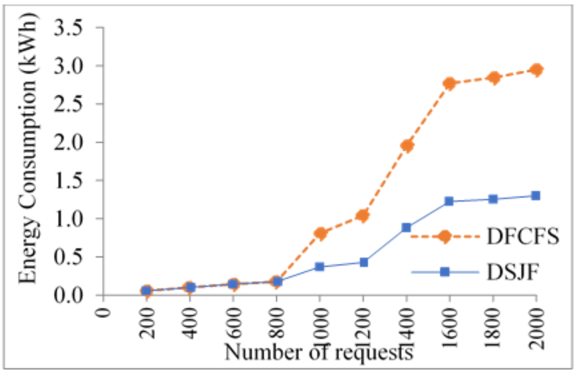

Figure 5. Energy consumption comparison of DFCFS and DSJF

7. Conclusion

Cloud data centers are the digital age factories, and data center power consumption becomes a global issue. The proposed algorithms are intended to save cloud data center energy consumption. To achieve this goal, under FCFS and SJF allocation policies, different power management techniques and different power models are compared and analyzed. The results proved that DVFS is saving more power than the other two power management techniques. The findings of the analysis showed that the cubic power model achieves greater energy efficiency and consumes less power than the other three power models. The evaluations for the proposed algorithms: DFCFS and DSJF are experimented using real workload traces of virtualized environment on Cloudsim simulator. The combination of SJF resource allocation algorithm and DVFS power management techniques with Cubic power model, called DSJF is very suitable for the case where run times of incoming requests are different from each other. The experimental results show DSJF algorithm gains higher energy efficiency and it can save up to 55% of energy consumption compared to DFCFS algorithm.

References

[1] T. Brey, L. Lamers, “Using virtualization to improve data center efficiency”, The Green Grid, Whitepaper, 2009.

[2] A. Varasteh, M. Goudarzi, “Server consolidation techniques in virtualized data centers: a survey”, IEEE Systems Journal, 11 (2), 2017, pp. 772-783.

[3] R. N. Calheiros, R. Ranjan, A. Beloglazov, C. A. De Rose, R. Buyya, “Cloudsim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms”, Software: Practice and Experience, 2011, pp. 23–50.

[4] A. Ali, L. Lu, Y. Zhu, J. Yu, “An energy efficient algorithm for virtual machine allocation in cloud data centers”, Springer Science and Business Media Singapore, J. Wu and L. Li (Eds.): ACA 2016, pp. 61–72.

[5] T. Yeboah, I. Odabi, K. K. Hiran, "An Integration of Round Robin with Shortest Job First Algorithm for Cloud Computing Environment", International Conference On Management, Communication and Technology, III.1, 2015, pp. 1–5.

[6] A. Beloglazov, J. Abawajy, R. Buyya, “Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing”, Future generation computer systems, 28(5), 2012, pp. 755-768.

[7] M. J. Usman, A. Samad, H. Chizari, & A. Aliyu, “Energy-Efficient virtual machine allocation technique using interior search algorithm for cloud datacenter”, 6th IEEE ICT International Student Project Conference (ICT-ISPC), 2017, pp. 1-4.

[8] S. R. Jena, V. Vijayaraja, A. K. Sahu, "Performance Evaluation of Energy Efficient Power Models for Digital Cloud", Indian Journal of Science and Technology, 9(48), 2016, pp. 1-7.

[9] L. Lee, K. Liu, H. Huang, C. Tseng, “A dynamic resource management with energy saving mechanism for supporting cloud computing”, Int. J. Grid Distrib. Comput. 6 (1), 2013, pp. 67–76.

[10] X. Fan, W. D. Weber, L. A. Barroso, “Power provisioning for a warehouse-sized computer”, Proc. 34th Annual Int. Sympo. Comput. Archit. (ISCA), ACM New York, USA, 2007, pp. 13–23.

[11] Kliazovich, D. Bouvry, P. Audzevich, Y. Khan, “Greencloud: A packet-level simulator of energy-aware cloud computing data centers”, IEEE International Global Telecommunications Conference, 2010, pp. 1–5.

[12] A. T. Makaratzis, K. M. Giannoutakis, D. Tzovaras, “Energy modeling in cloud simulation frameworks”, J. Futu. Gen. Comput. Syst., 79 , 2017, pp. 715–725.

[13] Parallel Workloads Archive. [Online]. Available: http://www.cs.huji.ac.il/labs/parallel/workload/