

Trends towards Energy Efficient with Backfilling based Scheduling Techniques for Cloud Computing

Amanjot Kaur
Department of Computer
Engineering & Technology,
Guru Nanak Dev University,
Amritsar, Punjab

Anil Kumar
Department of Computer
Engineering & Technology,
Guru Nanak Dev University,
Amritsar, Punjab

ABSTRACT

The OpenNebula is an open source platform which provides flexible and feature-rich cloud management solutions; one of them is Haizea which act as lease manager and it reinforce distinctive types of leases. The deadline sensitive lease is one of the supporting leases provided by Haizea. In real time scenario, the majority of the leases are deadline sensitive and these leases are scheduled by implementing the backfilling algorithm. The backfilling algorithm optimizes scheduling by sorting one of the leases from best effort queue and allocate the free resources to schedule the deadline sensitive lease. But in some cases, if there is the same kind of leases and should be in connective in sequence than backfilling algorithm does not provide an efficient platform for scheduling. AHP (Analytic Hierarchy Process) is used to enhance the backfilling algorithm, which acts as a decision maker in the backfilling algorithm to choose the possible best lease from the given best effort queue in order to schedule the deadline sensitive lease. The overall objective of this paper is to explore gaps associated with existing backfilling based scheduling techniques.

Keywords

Haizea, Backfilling, Scheduling, AHP, live migration, load balancing, energy efficiency, Open Nebula, Deadline sensitive.

1. INTRODUCTION

Now a day ubiquitous computing is emerging everywhere and cloud computing act as a model and also provide a shared pool of configurable computing resources [1]. Service provides allocates resources to the users on pay per use basis. The main challenge for cloud providers is how to schedule the resources in order to accomplish the unpredictable demands of the cloud users concurrently. During workload increases, it is very necessary for the service providers to manage the cloud resources so that performance of the cloud applications cannot be degraded [2]. To manage cloud resources, Open Nebula is used that is open source software toolkit. The responsibility of cloud service provider is to deliver on-demand resources to users. So the resource management is the critical aspect for the cloud service providers for better utilization of physical machines to enhance the server performance when workload increases.

1.1 Scheduling with Haizea

The Haizea is an open-source VM resource management component that accepts user requests and schedules them according to the availability of resources in the physical machines. Haizea is a kind of scheduler that acts a resource manager in the backend of OpenNebula [3]. It provides

resources to leases in the form of virtual machines that are implemented in cloud server data centers. The Haizea categorizes user requests falls into four types [7]:

Best-effort leases: These leases can be pre-empted to provide the resources requested by the Advance reservation leases and Immediate leases. Until the resources get free, the best effort leases will wait in a queue [3]. These leases do not have time limits associated with it.

Advance reservation leases: The AR lease starts at a particular time which requires resources within a strictly determined time period. These leases cannot be pre-empted and require time limits such as starting time and ending time.

Immediate leases: In this lease, the scheduler allocates the resources to these leases if resources are available right now, otherwise not. Immediate leases are also non pre-emptible in nature.

Deadline sensitive: To complete the request for the users who need resources in a specified time period for their best effort leases, completion time (deadlines) must be linked with the best effort leases. These type of leases are known as deadline sensitive that has particular starting time, ending Time (deadline), duration. Scheduling of deadline sensitive leases can be made possible by swapping and backfilling algorithms that are used in Haizea.

1.2 Swapping

Swapping can be done between two consecutive leases when the resources required by the first lease are smaller than the next lease and after swapping, leases should complete their execution within the specified time. We can provide a number of idle resources to a new lease by scheduling leases which require few numbers of requested resources behind the lease that needs a large number of requested resources [7].

Example1: This example illustrates how swapping of two successive leases leads to fulfill requests for resources required by new deadline lease [7].

Table 1 indicates the information of leases that submit its requests. It comprises of requested nodes, submit time of lease, start time of lease, duration in minutes and deadline (pm).

Table 1. Submitted leases

Leas e no.	Nod es	Submit time(A M)	Start time(P M)	Durati on	Deadline(p m)
1	2	11:10	12:00	40	01:20
2	3	11:20	12:00	20	01:20

3	4	11:30	12:00	20	01:20
4	1	11:40	12:30	40	01:20
5	2	11:50	12:40	10	01:20

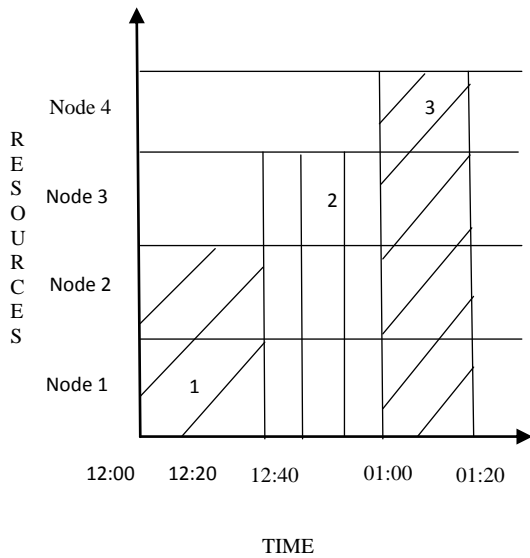


Fig 1: Existing algorithm

As shown in Fig. 1 the existing algorithm indicates that haizea will allocate nodes to leases according to their arrival time [7]. The lease which arrives earlier will be executed first. Therefore, the existing algorithm tries to schedule lease-1 first from (12.00 to 12.40) based on their arrival time as it comes first, then it processes lease 2 from (12.40 to 01.00) and then it processes lease-3 from (01.00 to 01.20). Now as it is shown that lease-4 requires 1 node from (12:30 PM to 01:20 PM) and already resources are occupied between this time period. So we cannot preempt lease-3 therefore due to lack of resources lease-4 is rejected and similarly lease-5 will also be rejected. So only 3 leases are scheduled by existing algorithm [7].

It has clearly mentioned that there are some idle nodes from 12:00 PM to 01:00 PM. So these idle nodes can be allocated to the new leases which will be possible by swapping in such a way that if two leases are requested for resources, the first lease will be that one which requires more resources and second lease requires fewer resources. So this process results in the non-ascending order of the resources which are requested by the leases.

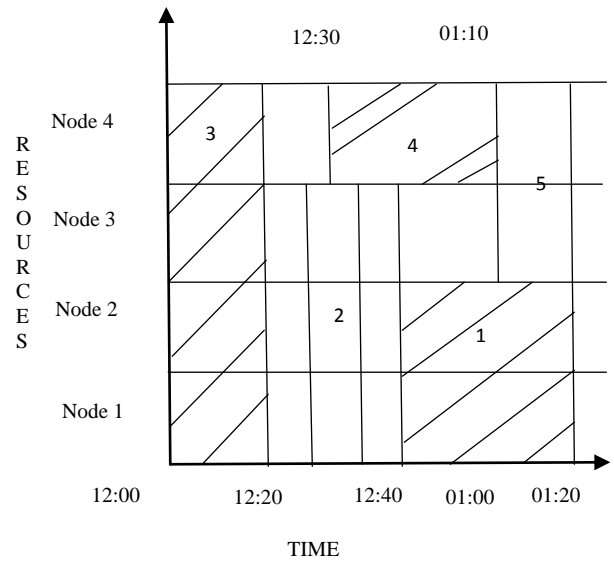


Fig 2: Swapping

Fig. 2 indicates how idle nodes are used to fulfill the request of new leases by using the swapping technique. Firstly, the scheduling of Lease-1, Lease-2 and Lease-3 will be done as shown in fig1 but the arrival of lease 4 leads to the again scheduling of preceding three leases. So, these leases are organized by swapping in the non-ascending order of expected nodes and hence lease-3 is rescheduled first then lease-2 and finally lease-1 as shown in Fig 2. Remaining resources is used to schedule lease-4 between 12.30 to 01.10 PM. Now lease-5 is scheduled easily as shown in Fig. 2 [7].

1.3 Backfilling

In backfilling, a decision maker (AHP) is used that selects the best effort lease from the best effort queue in order to provide idle resources in the free slots where a deadline sensitive lease can be scheduled [7]. The backfilling algorithm is based upon the slack value of the lease. The Slack value is calculated to verify, where the lease can be scheduled or not using the backfilling algorithm. This value of the lease indicates whether the lease is considered as a deadline sensitive lease or an advance reservation lease. It determines the preempt ability of the lease. The deadline, start time and duration of the lease are used as the parameters to calculate the slack value (1) of a lease [26].

$$\text{Slack Value} = \text{deadline} - \text{start time}/\text{Duration (1)}$$

If the calculated slack value of a lease is greater than the system administrator decided lower bound slack value, then the lease is considered as deadline sensitive lease [26].

Table 2. Submitted leases information

Lease No.	No des	Submit Time	Start Time	Durati on	Deadline(PM)
1	2	11:10	12:00	20	12:30
2	3	11:20	12:00	40	01:00
3	2	11:30	12:00	50	01:50
4	4	11:40	01:00	20	01:50

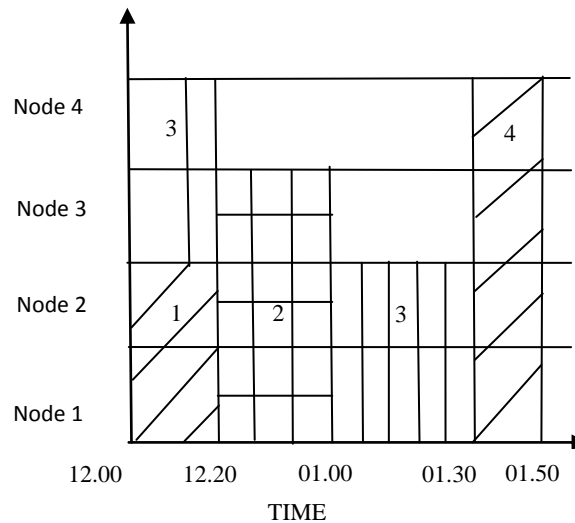


Fig 3: Backfilling

Example 2: This case succinct about backfilling of leases that can create space for a recently arrived deadline sensitive lease.

Table 2 depicts the information of requested resources, lodged leases, start time, Submit time of leases, duration in minutes and deadline [7].

As shown in Fig.3, All the leases are scheduled by using the backfilling algorithm. Lease-1 requires two nodes which is scheduled in the slot (12:00 PM to 12:20 PM) and results in two unused nodes in this slot. backfilling can be implemented to use these idle resources in the time slot (12:00 PM to 12:20 PM). As lease-2 cannot be executed at its start time, then these idle nodes are used by some other lease which are residing behind lease-2 in the collection of given leases. Now Lease-3 is executed in the time from 12:00 PM to 12:20 PM and is stopped at 12:20 PM to permit lease-2 to execute. The rest of the 30 minutes of lease-3 must be executed between 01:00 PM to 01:30 PM. So Backfilling results in 4 idle nodes in the slot (01:30 PM to 01:50 PM) which can be used to schedule lease-4. There is one limitation of backfilling is that it requires pre-emptions [7].

1.4 The analytic hierarchy process(AHP)

The analytic hierarchy process (AHP) is a decision-maker method which enables us to make effective decisions on complex issues [4]. The scheduling performance of the backfilling algorithm is improved by using AHP that is used as a decision maker to finds the lease from the best effort queue which provides free slots for allocating deadline sensitive lease [26]. The various Steps in AHP are as follows:

Hierarchy Development: The initial phase in the AHP is to build up a graphical representation of the problem regarding the general objective, the criteria, and the decision choices. For scheduling the deadline sensitive lease, the parameters such as deadline, duration and start time are taken as criterions for the AHP the same type of leases are considered as alternatives of the AHP [26].

Pairwise comparisons: Pairwise comparisons of elements in each group are computed and represented by a matrix. Matrix indicates the relative importance the criterions. Then, these values are calculated to know the ranks among the criterions. According to these rank values, the decisions are taken [4].

1.5 Issues related with Scheduling

There are three issues related with Scheduling [6].

Live Migration: Live migration is a technique of moving the running virtual machine from one physical host to other without disrupting the applications which are running on the machine. The advantages of Live migration include low energy consumption, load balance among cloud servers, fault tolerance in case of server failure.

Load Balancing: The main goal of load balancing is the distribution of load among the various physical servers which results in improving resource utilization, increasing throughput, minimizing response time, and preventing overloading of any single resources. The whole system load can be balanced by migrating virtual machines from heavily loaded physical machines to light loaded physical machines [25].

Energy efficiency: The energy consumption of cloud data centers depends on the server utilization level. Consolidation of virtual machines is performed to conserve the energy. The virtual machines which are underutilized are switched down to ensure energy efficient green cloud.

2. RELATED WORK

Suwendu Chandan Nayak, et al. [26] introduces AHP (Analytic Hierarchy Process) in the backfilling algorithm as a decision maker to find the best lease from the best effort queue in order to schedule the deadline sensitive lease to decrease the lease rejection and improves the performance of the backfilling algorithm using AHP.

Aarti Singh, et al. [14] implements an algorithm (A2LB) Autonomous agent based mechanism which concentrates on load adjusting in distributed computing environment.

Nidhi Jain Kansal, et al. [21] proposed a technique called Firefly optimization to enhance the energy efficiency of cloud data centers by live migration of virtual machines that moves the overloaded Virtual machines and hence maintaining the utilization of resources.

Fahimeh Farahnakian, et al. [17] introduces Ant-colony System which reduces the energy utilization rate of cloud servers by decreases number of active nodes. It tries to get a

best optimal solution and beats existing VM combination approaches in terms of performance of data center, energy utilization rate and a number of VM migrations.

Amit Nathani, et al. [7] proposed scheduling based dynamic planning algorithm which discovers numerous time slots for the allocation of resources required by deadline sensitive leases. Along with pre-emption it also describes the concept of swapping and backfilling during rescheduling of existing lodged leases to create a vacancy for a recently appeared lease and maximizes acceptance of leases and resource utilization.

Keng-Mao Cho, et al. [8] depicts optimization of ant colony including particle swarm (ACOPS) which helps for load balance in a dynamic environment. This technique tries to fulfill customized Virtual machine demands in which the resources such as memory, CPU and utilization of disk are seen for load balancing.

Hongjian Li, et al. [18] shows The Modified Particle Swarm Optimization algorithm (MPSO) which reduces the quantity of dynamic physical hubs, the measure of virtual machine movements and gives better energy proficiency in cloud centers. . It provides analysis among the conventional MBFD and MPSO algorithm.

ZHOU Zhou, et al. [15] proposed an algorithm called TESA to enhance the energy consumption of cloud servers. TESA allows migration of Virtual machines from over utilized nodes to the nodes that possess less load by maintaining appropriate load.

Aminu Abdulkadir Mahmoud, et al. [10] introduces an RMK technique in which for the first stage the, finish time of the shortest job is calculated which is based upon the finish time of the strategies that are min-min, suffrage, max-min and in next stage, genetic algorithm is applied to improve scheduling of the jobs and efficient load balancing of the resources.

Nasser R. Sabar, et al. [22] represents grammatical evolution dependent Solution to improve load balancing optimization

issue. Degang Sun, et al. [19] shows the comparison of the established methods for the protection of live migration and implemented a security design of migration based on SPLM (Security Protection of Live Migration) and investigate its security and consistency, which has demonstrated that SPLM is superior to anything others.

Flavio Esposito, et al. [23] presents a design for live migration as well as programming plan which leads to overall migration time to be minimum and proposed an algorithm, called GeoMig which improves current live-migration implementations.

Elham Akbari, et al. [24] shows the method called weighted linear prediction for improving the energy consumption rate of the cloud servers which aims to emulate the consequences of different workloads on energy utilization of the cloud framework. It describes the existing approach called MQMPM which is used to estimate the future usage of virtual machines.

Pradip D. Patel, et al. [6] represents a survey of the different techniques that are used for live migration. The present method called pre-copy paradigm for live migration are described. Many techniques that are explained decreases the downtime and enhances efficiency in low bandwidth.

Thomas L. Saaty, et al. [4] shows the Analytic Hierarchy Process (AHP) that acts as a decision-making tool which is applied to measure inconsistency in the judgements. AHP enhances the judgements by making pairwise comparisons.

3. COMPARISON TABLE

Table 3 is clearly representing various issues associated with scheduling techniques. It has demonstrated that no technique is effective in every case. Thus scheduling is still an open area of research. However, the AHP based scheduling has better results than majority of existing techniques. Therefore, in near future the live migration based AHP scheduling technique will be proposed for cloud computing environment in order to enhance the energy consumption rate further.

Table 3. Comparative analysis

Ref No	Year	Techniques	Load balancing	Energy efficiency	Live Migration	Backfilling
[7]	2012	dynamic planning based scheduling algorithm, swapping	✓	✗	✗	✓
[8]	2014	ACOPS(ant colony optimization with particle swarm)	✓	✗	✗	✗
[9]	2014	Markov overload detection (MHOD) algorithm, THR, LRR,MMTMC	✗	✓	✓	✗
[10]	2015	Multi-Criteria algorithm (RMK),genetic algorithm, Min-Min , Max-Min, Suffrage algorithm.	✓	✗	✗	✗
[11]	2015	Load Aware Virtual Machine Placement (LAVMP) algorithm	✓	✓	✓	✗
[12]	2015	Replication algorithm	✗	✓	✗	✗

[13]	2015	performance-related, Power-related and network traffic-related metrics.	×	✓	×	×
[14]	2015	Autonomous agent based load balancing algorithm (A2LB)	✓	×	×	×
[15]	2015	three-threshold energy saving algorithm (TESA)	✓	✓	✓	×
[16]	2015	improved redundant power consumption laxity-based (EIRPCLB) algorithm	✓	✓	×	×
[17]	2015	Ant Colony System (ACS) based VM Consolidation (ACSVMC)	✓	✓	✓	×
[18]	2016	Modified PSO algorithm (MPSO), MBFD algorithm.	✓	✓	✓	×
[19]	2016	SPLM(security protection of live migration)	✓	×	✓	×
[20]	2016	weighted random task scheduling algorithm	✓	×	×	×
[21]	2016	FireFly Optimization , FsFO-EVMM	✓	✓	✓	×
[22]	2016	GE enhanced SA method (GE-SA)	✓	×	×	×
[23]	2016	GeoMig algorithm	×	×	✓	×
[24]	2016	MQMPM, MQMPM/WLPM	×	✓	×	×
[26]	2016	AMLB algorithm, novel load balancing algorithm	✓	×	×	×
[27]	2016	AHP (Analytic Hierarchy Process)	✓	×	×	✓

4. CONCLUSION AND FUTURE WORK

Cloud computing has become very popular due to its wide range of applications via the internet. The service composition based techniques that are conscious from the server selection from the cloud can progress to the cost and efficiency of cloud computing. In the above comparison table, it represents the various no of parameters i.e. load balancing, energy efficiency, live migration and backfilling which has shown that the backfilling algorithm has been enhanced by AHP where energy efficiency and live migration has not been considered in existing research. So in near future work, we will evaluate the live migration and energy efficiency for cloud computing environment in order to enhance the energy consumption rate further.

5. REFERENCES

- [1] Mell, Peter, and Tim Grance. "The NIST definition of cloud computing." (2011).
- [2] Liu, Jinzhao, et al. "Aggressive resource provisioning for ensuring QoS in virtualized environments." *IEEE Transactions on Cloud Computing* 3.2 (2015): 119-131.
- [3] Haizea.cs.uchicago.edu, 2016. Haizea – An Open Source VM-Based.
- [4] Lease Manager". N.p., 2016. Web. 15 Jan. 2016.Saaty, Thomas L. "Decision making with the analytic hierarchy process." *International journal of services sciences* 1.1 (2008): 83-98.
- [5] Saaty, Thomas L. "Decision-making with the AHP: Why is the principal eigenvector necessary." *European journal of operational research* 145.1 (2003): 85-91.
- [6] Patel, Pradip D., et al. "Live Virtual Machine Migration Techniques in Cloud Computing: A Survey." *International Journal of Computer Applications* 86.16 (2014).
- [7] Nathani, Amit, Sanjay Chaudhary, and Gaurav Somani. "Policy based resource allocation in IaaS cloud." *Future Generation Computer Systems* 28.1 (2012): 94-103.
- [8] Cho, Keng-Mao, et al. "A hybrid meta-heuristic algorithm for VM scheduling with load balancing in cloud computing." *Neural Computing and Applications* 26.6 (2015): 1297-1309.
- [9] Beloglazov, Anton, and Rajkumar Buyya. "OpenStack Neat: a framework for dynamic and energy-efficient consolidation of virtual machines in OpenStack clouds." *Concurrency and Computation: Practice and Experience* 27.5 (2015): 1310-1333.
- [10] Mahmoud, Aminu Abdulkadir, et al. "Multi-Criteria Strategy for Job Scheduling and Resource Load

- Balancing in Cloud Computing Environment." *Indian Journal of Science and Technology* 8.30 (2015).
- [11] Liu, Chengjiang. "A Load Balancing Aware Virtual Machine Live Migration Algorithm." (2016).
- [12] Boru, Dejene, et al. "Energy-efficient data replication in cloud computing datacenters." *Cluster Computing* 18.1 (2015): 385-402.
- [13] Fiandrino, Claudio, et al. "Performance and energy efficiency metrics for communication systems of cloud computing data centers." (2015).
- [14] Singh, Aarti, Dimple Juneja, and Manisha Malhotra. "Autonomous agent based load balancing algorithm in cloud computing." *Procedia Computer Science* 45 (2015): 832-841.
- [15] Zhou, Zhou, et al. "A novel virtual machine deployment algorithm with energy efficiency in cloud computing." *Journal of Central South University* 22 (2015): 974-983.
- [16] Enokido, Tomoya, Dilawaer Duolikun, and Makoto Takizawa. "An extended improved redundant power consumption laxity-based (EIRPCLB) algorithm for energy efficient server cluster systems." *World Wide Web* 18.6 (2015): 1603-1629.
- [17] Farahnakian, Fahimeh, et al. "Using ant colony system to consolidate vms for green cloud computing." *IEEE Transactions on Services Computing* 8.2 (2015): 187-198.
- [18] Li, Hongjian, et al. "Energy-efficient migration and consolidation algorithm of virtual machines in data centers for cloud computing." *Computing* 98.3 (2016): 303-317.
- [19] Sun, Degang, et al. "SPLM: Security Protection of Live Virtual Machine Migration in Cloud Computing." *Proceedings of the 4th ACM International Workshop on Security in Cloud Computing*. ACM, 2016.
- [20] Qian, Zhang, et al. "A Load Balancing Task Scheduling Algorithm based on Feedback Mechanism for Cloud Computing." *International Journal of Grid and Distributed Computing* 9.4 (2016): 41-52.
- [21] Kansal, Nidhi Jain, and Inderveer Chana. "Energy-aware Virtual Machine Migration for Cloud Computing-A Firefly Optimization Approach." *Journal of Grid Computing* 14.2 (2016): 327-345.
- [22] Sabar, Nasser R., and Andy Song. "Grammatical Evolution Enhancing Simulated Annealing for the Load Balancing Problem in Cloud Computing." *Proceedings of the 2016 on Genetic and Evolutionary Computation Conference*. ACM, 2016.
- [23] Esposito, Flavio, and Walter Cerroni. "GeoMig: Online Multiple VM Live Migration." *Cloud Engineering Workshop (IC2EW), 2016 IEEE International Conference on*. IEEE, 2016.
- [24] Akbari, Elham, et al. "Incorporation of weighted linear prediction technique and M/M/1 Queuing Theory for improving energy efficiency of Cloud computing datacenters." *Long Island Systems, Applications and Technology Conference (LISAT), 2016 IEEE*. IEEE, 2016.
- [25] Chien, Nguyen Khac, Nguyen Hong Son, and Ho Dac Loc. "Load balancing algorithm based on estimating finish time of services in cloud computing." *2016 18th International Conference on Advanced Communication Technology (ICACT)*. IEEE, 2016.
- [26] Tripathy, Nayak Chitaranjan. "Deadline Sensitive Lease Scheduling in Cloud Computing Environment Using AHP." *Journal of King Saud University-Computer and Information Sciences* (2016).