Survey of Task Scheduling in Cloud Computing

Arezoo Abbasi*, Ramin Karimi

Department of Computer Engineering, Islamic Azad University, Tehran,

Iran

Review Article

*For Correspondence : Arezoo Abbasi, Department of Computer Engineering, Islamic Azad University, Tehran, Iran, Tel: 9809334309401; Email: arezoo_abbasi24@yahoo.com

ABSTRACT

Cloud computing, the Internet based computing that won a lot of momentum for its flexibility and elasticity, provides shared data and processing resources as services allowing users to not need to have their own infrastructure and follows the pay-as-you-go model. A Cloud computing system based on virtualization technology can enhance the flexibility of load deployment through dynamic resource expansion. The scheduling is defined into two levels of host and Virtual Machine (VM). In virtual machine level, the tasks are allocated to virtual machines, and in host level, the virtual machines to the hosts. The optimization of tasks performance is one the most important subjects in cloud environments. In this paper, we review several aspects of the scheduling literature for cloud computing.

INTRODUCTION

The collection of interconnected computers that consists of more than one united calculation resource is known as the cloud. In recent years, the advancement of cloud computing has helped simulate the quick arrangement cloud computing

of inter connected data centers that are geographically dispersed for offering high quality and dependable services. These days, cloud computing has turned into an efficient paradigm to offer computational abilities on a ''payper-utilize'' premise. Cloud computing brings the adaptation and change in the IT business. With its developing application and promotion, cloud computing offers wonderful open doors, as well as confronts many difficulties in the advancement of traditional IT recently, cloud computing has risen as other Internet based model for empowering clients. It can organize availability to a shared pool of configurable assets on request, which can be immediately provided and discharged with very little administration or cloud provider collaboration because of this novation, many advantages such as improved benefits in the marketplace place with respect to time, cost, stack adjusting, and storage can be realized. A scheduling algorithm is utilized to plan the task with greatest evaluated gain or benefit and execute the task. However, computing ability in the distributed system shifts from various resources to the cost of resource utilized. These tasks are scheduled in view of client's necessity. Additionally, if the number of clients using the cloud increases, scheduling becomes quite difficult and an appropriate scheduling algorithm needs to be utilized. In the early stage, some of the scheduling algorithms were developed in the context of grid computing and based on their performance many were adapted for distributed computing. In cloud computing, users may utilize hundreds or thousands of virtualized resources and it is impossible for everyone to allocate each task manually. Due to commercialization and virtualization, cloud computing handles the task scheduling complexity in the virtual machine layer. Thus, scheduling plays an important role in cloud computing to assign the resources to each task efficiently and effectively [1-3].

LITERATURE REVIEW

Related work

An energy conscious scheduling algorithm with a load balancing approach is proposed by that uses bee colony feed search behavior to detect high and low load hosts. Virtual machines migrate from high load hosts to low load hosts. Also, if the host has a very low load, all its machines are migrated to other hosts to shut down the host in order to reduce energy consumption [4]. This algorithm measures the level of Service Level Contract (SLAV) breach to evaluate the quality of its work. The proposed algorithm significantly reduces energy consumption compared to conventional methods and similar energy aware methods [5].

The authors in proposed a scheduling algorithm to increase load balance and reduce energy consumption. In the standard ant colony method, ants first build their own solution and then produce a general solution, but in the ants are constantly updating a single set of answers instead of updating individual responses. Compared to the standard ant colony, this article has succeeded in reducing the rate of service level contract violations, reducing energy consumption, and increasing load balance.

Energy allocation, scheduling and efficiency are important issues in cloud systems. In this article, allocation and scheduling to minimize the number of active servers. In this paper, feverfew algorithm is used for the allocation process. This algorithm belongs to the group of accidental algorithms, which means that a kind of random search is used to reach a set of solutions [6]. The main characteristic of fireflies is their flashing light. Fireflies move towards a worm that is brighter. The simulation results show that the proposed method has improved at runtime. One of the disadvantages of this article is that it does not take into account the criteria of energy and improving resource efficiency.

An efficient binary bee colony based algorithm for scheduling tasks is presented by article, which uses a flexible ranking and scoring policy in the production stages of the initial colony population. The purpose of using this technique is to

increase the speed of convergence in future generations and to facilitate exploration in the space of global problem solving. The main purpose of this algorithm is to reduce the completion time. Compared to the standard bee colony, it has decreased the amount of completion time [7]. This research has a comprehensive evaluation but only considers the criterion of completion time.

In this paper a load balancing method using high and low thresholds for selecting physical machines for virtual machine migration is proposed. The selected virtual machines are then placed in a host that consumes the least amount of energy to minimize energy consumption [8]. To evaluate the proposed method is compared with recent load balancing methods. An effective load balancing approach can reduce migration and energy consumption by minimizing the number of servers. In this paper, a load balancing method that uses the benefit of a physical host to migrate virtual machines is proposed. The proposed method selects the virtual machine when the physical host is overloaded or under loaded and migrates it to the host that consumes less energy. This paper proposes an algorithm for selecting and scheduling a virtual machine. The simulation results show the reduction of energy consumption, migration and simulation time.

In presents a scheduling algorithm based on the genetic algorithm method, which tries to reduce response time and execution time by focusing on user satisfaction and achieving service quality [9]. User requests are divided into four categories based on time and budget constraints. To improve the results of the genetic method, a restart process has been added to the population update phase. Thus, if the best fit of the current population is below the threshold, half of the best chromosomes of the current population are added to the new population and the rest of the chromosomes of the new population are randomly generated.

A good task scheduler should adjust its scheduling technique to the changing condition and the kinds of task. Consequently, a dynamic task scheduling calculation, for example, Ant Colony Optimization (ACO), is fitting for clouds. ACO calculation is an arbitrary search algorithm. This algorithm utilizes a positive criticism system and mimics the conduct of genuine ant colonies in nature to scan for food and to joint with each other by pheromone laid on ways voyaged. Numerous scientists utilized ACO to take care of NP-difficult issues, for example, voyaging sales representative issue, graph coloring issue, vehicle routing issue and scheduling issue. Earlier, a lot of researchers used this algorithm to explain the scheduling approach. Mathiyalagan, et al. used an enhanced ant colony algorithm to describe grid scheduling. To develop the competence of the system, they further brought in an adapted ant colony optimization algorithm for programming in grid computing, as exemplified in the work of Mathiyalagan, et al [10].

In this paper, a dynamic energy driven virtual machine scheduling and migration algorithm based on a multi-source model is developed. This model can reduce energy consumption while maintaining service quality. In this model, a two threshold method using several sources for virtual machine migration is presented [11]. Modified Particle Swarm Optimization (MPSO) method has been introduced for aggregation and scheduling of virtual machines. The main contributions of this article include:

- The multi-source energy-based aggregation and migration algorithm of virtual machines under dynamic load has been designed and implemented.
- A two threshold method based on the use of multiple sources for virtual machine migration and the MPSO algorithm for redistribution of virtual machines is expressed.
- Analysis between traditional MBFD and MPSO heuristic algorithms

Compared to the MBFD algorithm, it has reduced the number of physically active nodes and the number of migrations. Clusim simulator is used to simulate the migration of virtual machines with dynamic multi-source load. The results show an improvement in energy efficiency, a reduction in the number of physically active nodes, and the migration of virtual
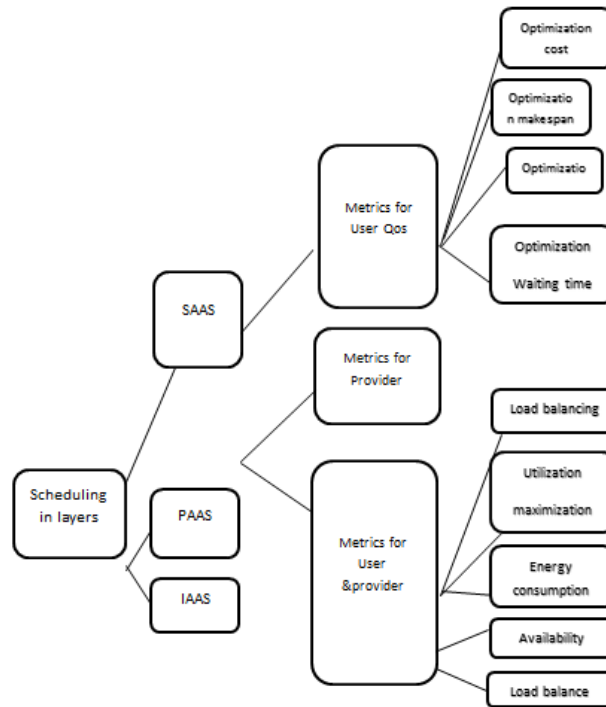
machines [12].

In this paper, using the completion time (Makespan) and cost, a solution is proposed and these two parameters are considered as main constraints for the optimization problem. In this paper, we merge two algorithms called Cuckoo Search Algorithm (CSA) and Opposition Based Learning (OBL) and combine a new hybrid algorithm called the Cuckoo Search Algorithm (OCSA) to create solutions to the above. Has provided. The proposed OCSA algorithm has shown a significant improvement over other task scheduling algorithms. The main purpose of the proposed way is to schedule tasks with the minimum completion time and cost. In this article, scheduling is done in parallel. For example, all tasks are processed simultaneously. Scheduling plays an important role in navigating tasks within the cloud. The scheduling process first analyzes how much resources are needed to complete the tasks and which task should be allocated to which computational component. Scheduling-based optimization method is proposed to overcome the problems in current scheduling methods to minimize cost and execution time.

Cloud resource layers

Based on resource distribution type, cloud computing is composed of three layers Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) and in this way services are offered for users. IaaS: This layer is in the lower level and supports the needed infrastructure for servers like processors, storage and hardware Equipment. So users do not need to buy server, data center and network resources anymore. In addition, costs are paid based on the duration of using them. On the whole, it is feasible to say resource management is done in this level; such as Amazon EC2 [13,14]. PaaS: This is the middle layer of needed software and supporting them. This layer is a hosting for different environments offering services which are provided package; such as Microsoft Azure, Google application engine. SaaS: This is related to online software, such as an online office or Google Doc, which resolves various problems of traditional software like updating and deployment. In addition, it is helpful for reducing costs. For instance, Sale force CRM is software on cloud. Scheduling can be built in different layers of the service heap. As you can see in Figure 1, in different layers cloud has its own strategy. Therefore, the architecture composed of the SaaS, PaaS, IaaS heaps can be used to categorize scheduling problems into software layer, platform layer and infrastructure layer. Scheduling in SaaS is to schedule VMs to task, job, user applications, etc.

Cloud resource layers based on resource distribution type, cloud computing is composed of three layers substruction as a service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) and in this way services are presented for users [14]. IaaS: This layer is in the lower level and supports the needed infrastructure for servers like processors, storage and hardware equipment. So users do not need to buy server, data center and network resources anymore. In addition, costs are paid based on the duration of using them. On the whole, it is possible to say resource management is done in this level; such as Amazon EC2 [5].

**Figure 1.** Scheduling in different layers.

DISCUSSION

Task scheduling

The scheduling is defined into two levels of host and virtual machine. In virtual machine level, the tasks are allocated to virtual machines, and in host level, the virtual machines to the hosts. The optimization of tasks performance is one the most important subjects in cloud environments. One of the problems is when a great number of users ask cloud resources at the same time; these difficulties could be solved by correct task scheduling to the available VMs. The aim of scheduling is usually loading distribution on VMs based on parameters like cost, response time, reliability, speed, utility, scheduling successful rate, deadline time, etc. Generally, in scheduling algorithms, there is a list of tasks attributed to all tasks made on the base of their priorities. The tasks are chosen based on their priorities and are allocated to the VMs [15].

System organization in distributed cloud systems

In distributed systems, the overall system organization may have influence in the scheduler scope. This organization can be a result of the substructure architecture or existing policies and access controls. Regarding the system management organization, it can be classified in three categories: centralized, decentralized, and hierarchical (Figure 2).

Figure 2. Taxonomy for scheduler's organization in distributed systems.
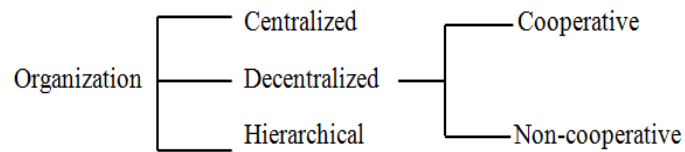
In figure, in a centralized management, the distributed system has a central entity that has information about the whole system and is accountable for managing its resources and jobs execution.

On the other hand, a decentralized management is done through distributed algorithms that manage the system without a central entity. In such systems, management roles may be distributed over resources in the system [4]. Scheduling in a decentralized system is created by more than one scheduler, where each scheduler has information about part of the system and may interchange information with other schedulers to make decisions in a cooperative way. A decentralized non-cooperative distributed system considers that each scheduler is responsible for achieving its own objectives, irrespective the objectives of other schedulers.

A hierarchical system structures the management in a tree like structure. Therefore, machines at a given level can communicate with machines at a level immediately above or below it in the hierarchy. With this, the distributed system management is shared, with entities that are higher in the hierarchy taking high level decisions while entities at the bottom make the low level decisions.

Note that the centralized architecture can be seen as a particular case of the decentralized model where only one scheduler exists. In addition, algorithms for the centralized architecture can independently execute in each scheduler of a decentralized (or hierarchical) system. In this case, a scheduler component could be relationship for the communication among different schedulers to arrange jobs execution within different boundaries. A more detailed view of distributed systems organization can be found in.

Dependent parameters in scheduling

To evaluate each algorithm or scheduling, different parameters and variables could be taken into account, such as:

Waiting time: the sum of elapsed time of a task in waiting line.

Response time: the time of response is calculated from the time of task entering to system until the time of its exiting.

- Make span: it is the subtraction time of finish time of the last VM from the start time of the first VM.
- Cost: the cost that each customer pays for the time use of VM.
- Throughput: the number of task done in the unit of the time.
- Utility: the duration of task execution in each of VM in proportion to time of total execution.
- Time execution: the duration when the users use VMs.
- Deadline: the time limit for performing the task and if this limit is not observed the Provider have to pay fine.
- CPU idle time: the duration when the CPU is workless and has no task to do. All of these methods and parameters are taken in algorithms to have a schedule in which the most utility be met in VMs [15, 16].

CONCLUSION

In this paper, we review, classify and evaluate task scheduling in cloud centers. We have shown that the most main and prominent goals in the field of research are the goals of completion time, cost. Therefore, as can be seen from the literature in this field, it is a challenge to provide scheduling algorithms that consider different objectives such as time, cost and energy together and consider multi objective optimization. As a result, providing an algorithm for scheduling tasks is an important issue that can be effective in reducing energy consumption and costs.

REFERENCES

1. Kaiyue W, et al. Distributed online scheduling and routing of multicast-oriented tasks for fitness function driven cloud computing. IEEE Commun Lett. 2018;20:684–687.

2. Zhu X, et al. ANGEL: Agent-based scheduling for realtime tasks in virtualized clouds. IEEE Trans Comput. 2018;64:3389–3403.

3. Keshanchi B, et al. An improved genetic algorithm for task scheduling in the cloud environments using the priority queues: formal verification, simulation, and statistical testing. J Syst Softw. 2017;124:1–21.

4. Ghafari SM, et al. "Bee-MMT: A load balancing method for power consumption management in cloud computing." In Contemporary Computing (IC3), 2013 Sixth International Conference on 08-10 August, Noida, India. 2019;76-80.

5. Khan S, et al. "Effective scheduling algorithm for load balancing (SALB) using ant colony optimization in cloud computing." Int J Adv Res Comput Sci Softw Eng, 2018;4:966-973.

6. Aruna M, et al. "An improved load balanced metaheuristic scheduling in cloud." Cluster Computing. 2019;22:10873-10881.

7. Kim SS, et al. "Optimal job scheduling in grid computing using efficient binary artificial bee colony optimization." Soft computing. 2019;17:867-882.

8. Jain, et al. "Enhanced load balancing approach to optimize the performance of the cloud service using virtual machine migration." Int J Eng Manuf (IJEM). 2018;7:41-48.

9. Jang SH, et al. "The study of genetic algorithm-based task scheduling for cloud computing." Int J Control Autom, 2019;5:157-162.

10. Yi, et al. "Task optimization and scheduling of distributed cyber–physical system based on improved ant colony algorithm." Future Gener Comput Syst. 109:134-148.

11. Hongjian LI, et al. "Energy-efficient migration and consolidation algorithm of virtual machines in data centers for cloud computing." Computing. 2018;98:303-317.

12. Krishnadoss, et al. "OCSA: task scheduling algorithm in cloud computing environment." Int J Intell Eng Syst. 2019;11:271-279.

13. Gao Y, et al. "Service level agreement based energy-efficient resource management in cloud data centers." Electr Comput Eng. 2018;40:1621-1633.

14. Beloglazov, et al. "A taxonomy and survey of energy-efficient data centers and cloud computing systems." Adv Comput. 2018;82:47-111.