Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 125 (2018) 725–732

Procedia Computer Science

www.elsevier.com/locate/procedia

6th International Conference on Smart Computing and Communications, ICSCC 2017, 7-8 December 2017, Kurukshetra, India

Modified HEFT Algorithm for Task Scheduling in Cloud Environment

Kalka Dubey[a*], Mohit Kumar[b], S.C. Sharma[a,b]

[a]*Research Scholar, IIT Roorkee 247667 India*
[b]*Research Scholar, IIT Roorkee 247667 India*
[a,b]*Professor, IIT Roorkee 247667 India*

Abstract

Cloud computing is now dominated in the area of high performance distributing computing and it provides resource polling and on demand services through internet. Therefore task scheduling becomes an important research area in the field of cloud environment because user's services demand change dynamically. Heterogeneous Earliest Finish Time (HEFT) unable to distribute the task efficiently. We modify HEFT algorithm that distribute the workload among the processor in effective way and reduce the makespan time of applications. Computational results (Fig. 4-5) shows that modify HEFT algorithm perform better than existing HEFT, Heterogeneous Earliest Finish Time (CPOP) algorithm.

Peer-review under responsibility of the scientific committee of the 6th International Conference on Smart Computing and Communications

Keywords: Cloud computing, NP hard, Task Scheduling, HEFT

1. Introduction

Cloud Computing model has appeared with the growth of internet and its services which provided by internets to its user. Cloud model is based on distributed computing and consisting of collection of various virtualized computers

* Corresponding author.
E-mail address: kalka.dubey267@gmail.com

1877-0509
Peer-review under responsibility of the scientific committee of the 6th International Conference on Smart Computing and Communications
10.1016/j.procs.2017.12.093

that can be dynamic in nature and inter connected to form computing resources. Globally vacant resources need to be utilized for increasing the utilization rate and earning from resources by increase the economic efficiency of these resources; cloud model is best suited for this purpose. The main aim of cloud computing model is sharing of resources and data to the users. It is a platform to provide the services and applications to its users. Cloud computing provided three types of Services software as a service (SaaS), platform as a service (PaaS) and infrastructure as a service (IaaS) [7]. These services is available to the users on the basis of pay per-Use-Demand, in which Shared computing resources, Servers, Data Storage, application and network. In SaaS service licensed of software is provided to the user on the basis of services subscription. These services can be access from any machine through the web browser. In PaaS user can create his own services with the use of available services of cloud and then deploy their services to own machine. In IaaS organizational Infrastructure is available to customer over the internet. Customer does not need to understand the internal architecture of infrastructure for using this. Instead of buying the whole infrastructure for business requirement customer take is as a rent basic when they required and when the requirement of infrastructure has no more the amount has paid for the services is used by the customer. In recent year the number of cloud users increased so the amount of tasks has need to manage propositionally for this task scheduling is required.

2. Task Scheduling

Task Scheduling is a technique of finding the order in which tasks or activity should be completed. It is mapping the resources to the appropriate task which is submitted for their completion to the cloud it's come in the category of NP hard problem because of large number of solution space and takes longer time for determine the optimal solution. It is a technique for management of resources in cloud. Task scheduling is solved the problem of which resources is to be allocated to which task so that increase the resource utilization and decrease the execution time. For a better performance scheduling algorithm need to be efficient and it consider load balancing of the overall system, interruption handling, fault tolerance, decrease the total execution time.

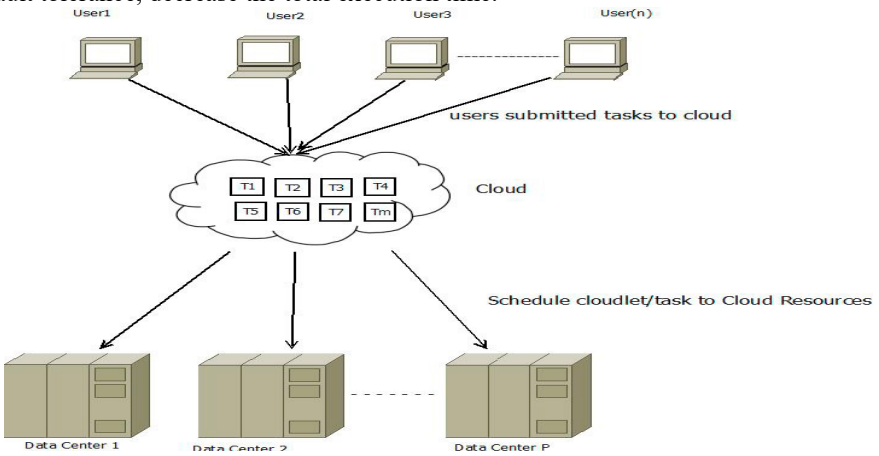

Figure 1 Scheduling in Cloud Computing

Users submitted their tasks for completion to cloud, these task need to assign to the processor for their execution. Now the concern is that how the tasks are assigned to processor so that minimum execution time and maximum profit is earned by the cloud owner. So here the task scheduling resolve the problem of assigning the tasks to the best suitable processor which considering the other factor. Task scheduling is best methods for better utilization of resource and achieving the economic efficiency. Various tasks scheduling method has been proposed and implemented in different scenario.

Based on the work of literature [2-3][6-13][16-17], we classified task scheduling methods in two groups in cloud environment. Distributed scheduling where the tasks assignment on different resources which are not located geographically on same place and the centralised scheduling where all the resources is on same place but the

complexity level is low compare to distributed scheduling . Distributed scheduling methods are further classified in three types heuristic, hybrids and metaheuristic methods. Heuristic methods classified into static and dynamic and hybrid methods are cost based, energy based, efficiency based and quality of service (QoS) based. Metaheuristic methods are nature inspired and swarm intelligence. Some static task scheduling algorithms are min-min, Symbiotic Organism Search (SOS), FUGE, HEFT and CPOP algorithms.

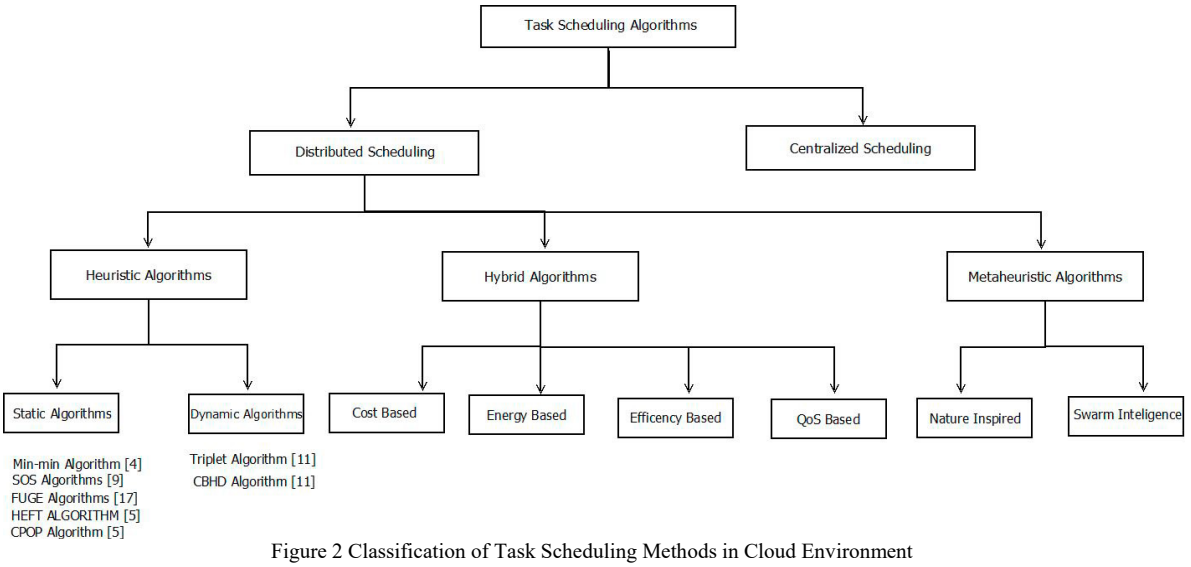

Figure 2 Classification of Task Scheduling Methods in Cloud Environment

3. Literature review

Table 1 Literature review on task scheduling

S.N	Year	Paper Title	Scheduling Parameters	Tool used to simulate	Improvement	Limitations
1	2016 [8]	"Task Scheduling Algorithms with Multiple Factor in Cloud Computing Environment"	cost and load balancing metrics	CloudSim	Improve the performance compare to the traditional algorithms.	Only compare with the traditional algorithms not compare any metaheuristic algorithm.
2	2016 [2]	"Symbiotic Organism Search optimization based task scheduling in cloud computing environment."	degree of Imbalance, makespan time and total execution time	CloudSim	Performance is improved by decreasing the degree of imbalance and makespan time	Do not consider task scheduling only focus on load balancing.
3	2016 [9]	"Enhanced Bee Colony Algorithm for Efficient	degree of imbalance,	Physical cloud environment	Decrease the makespan time	Compared only one algorithm

		Load Balancing and Scheduling in Cloud “	makespan time and total cost	using workstation	and improved the overall performance.	(Bee Colony)
4	2015 [10]	“FUGE: A joint meta-heuristic approach to cloud job scheduling algorithm using fuzzy theory and a genetic method”	degree of imbalance, makespan time, Execution time and cost	CloudSim	Execution cost and time is decreased and the overall performance is increased.	Energy consumption is highly affected factor which need not to be taken out.
5	2015 [3]	“Credit Based Scheduling Algorithm in Cloud Computing Environment”	makespan and priority of task	CloudSim	Resource utilization if increased and makspan time is decreased	Performance of proposed algorithm is not compared with the base algorithm.
6	2014 [11]	“Hybrid Job Scheduling Algorithm for Cloud Computing Environment”	degree of Imbalance and makespan time	CloudSim	Execution cost is decreased and overall profit is increased.	Only consider cost parameter do not take care other important parameter.
7	2014 [12]	“A task scheduling algorithm based on genetic algorithm and ant colony optimization in cloud computing”	Execution time	Cloud IM Simulator	Searching efficiency of Proposed algorithm has increased.	Take only few task need to be take more tasks so the result is more analyzed.
8	2014 [13]	“Improved Ant Colony Algorithm based on PSO and Its Application on Cloud Computing Resource Scheduling”	Execution time	MATLAB	Convergence is fast compared to the exiting algorithm.	Only consider the local minima need to take care of global minima as well.
9	2014 [7]	“Tasks scheduling technique using league championship algorithm for makespan minimization in IaaS cloud”	makespan time	MATLAB	makespan time has decreased and overall performance affected.	Comparison is not proper .
10	2014 [6]	“Task scheduling algorithm based on improved Min—Min algorithm in cloud computing environment”	makespan time and load balancing	CloudSim and GridSim	Performance is higher the exiting algorithm.	Only two parameter compared with Min-Min algorithm.
11	2012 [16]	“Dynamic task scheduling algorithm with load balancing for heterogeneous computing system”	makespan time, load balancing, resource utilization	Distributed Algorithm Simulator	Clustering is used to resolve the load balancing and minimize the makespan time.	Complexity has been increased.

12	2002 [17]	"Performance effective and low complicity task scheduling for heterogeneous computing "	schedule length ratio (SLR), speed up, execution time	Random graph generator	Decrease the total execution time and effective to achieve better schedule length ratio.	Suffer from the load balancing

4. Proposed Algorithm

Based on the above literature study we analyses various task scheduling algorithm, parameter, tools, improvement and limitation of algorithms. The main concept behind the HEFT algorithm is that a rank is calculated calculates for every task in DAG. HEFT algorithm work in two phases, in the first phase rank value is calculated while in the second phase processor assignment is done. HEFT algorithm start with set the weight of every node in DAG and set all the communication cost between nodes in the graph. For calculating the rank of task start with last node and move upwards in DAG until the root node has not came.

HEFT algorithm have the load balancing issue some of resources remain idle while other are overloaded so we proposed an algorithm which resolve some issue of HEFT algorithm. Based on the above literature study we analyses various task scheduling algorithm, parameter, tools, improvement and limitation of algorithms. The main concept behind the HEFT algorithm is that a rank is calculated calculates for every task in DAG.

Table 2 Modified HEFT algorithm

1. Create a DAG for all the submitted tasks T_i in Cloud.
2. Set the Computation Cost of tasks T_i and communication edges between the processor/resources R_j.
3. **For** i=1 to T_i (for every tasks in DAG calculate the order of execution, start with the last node in DAG)
4. **if** T_i is the last task then
5. Order of task OT_i = Average of task on all the processor
6. **else**
7. order of task OT_i = Average of task on all the processor + max (order of task value of predecessor task of current task) + communication cost between predecessor task node to current node
8. **endif**
9. **end for**
10. Arrange tasks in a list in decreasing manner on the basis of their order of task OT_i value.
11. **for** task in the list
12. map task to the processor which have the minimum execution time
13. **end for**
14. **End**

5. Result Analysis

In order to evaluate performance our proposed algorithm we take example of a DAG in figure 2 with communication cost between the nodes on three processor in table 3. First the order of task (OT) value is calculated of each task and make a list in which task is sorted in decreasing order of their OT value. We calculated the OT value of each task in table 3 by traversing the DAG in upwards direction starting from the exit node and traverse until the root node has not came. Arranged these tasks in a list by their OT value {T_1, T_2, T_5, T_3, T_4, T_6, T_7, T_8, T_{10}, T_9, T_{11}}. Second phase of the proposed algorithm is resource selection, select the best appropriate resource to task and map the task to that resource. Started with the first task in list that is T_1 compare the execution time of T_1 to

resource R_1, R_2, R_3 and find that resource R_1 is best so map the task to this resource. We had done mapping all tasks in figure 4.

Table 3 Computation time for each node on three processor

Task	Computation Time on R_1	Computation Time on R_2	Computation Time on R_3
T_1	16	19	27
T_2	18	15	13
T_3	21	12	22
T_4	15	13	11
T_5	22	19	20
T_6	13	09	11
T_7	8	11	16
T_8	14	23	10
T_9	28	32	12
T_{10}	15	13	09
T_{11}	14	16	22

As an illustration, figure 5 shows the makespan time of all three algorithm calculated by sample DAG of figure 2. We analyzed that makespan time of proposed algorithm is 133, is less than the makespan time of other two algorithms (HEFT and CPOP) and also take care load balancing issue. The advantages of better load balancing is that neither resource is overloaded nor under loaded, load is uniformly distributed but in HEFT and CPOP algorithm load is not uniformly distributed some of its resources are overloaded while other do not have any load, they are either in idle condition or have a very few load and this will increased the execution cost of task and affected the overall performance of cloud. Our proposed model tried to overcome the load balancing problem with better makespan time comparatively.

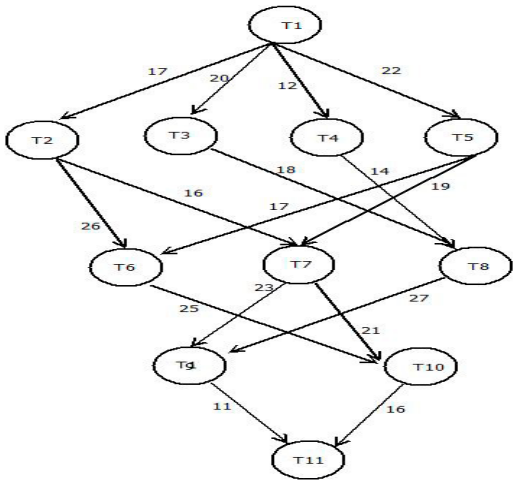

Figure 3 An application DAG example

Table 4 Calculation of order of task (OT) in upward direction from figure 3

Task	Average of task on all processor (A)	Max(predecessor) (B)	Communication Cost between predecessor and current node (C)	Order of task OT= (A+B+C)
T_1	20.666	134.666	17	172.332
T_2	15.333	93.333	26	134.66
T_3	18.333	83.329	18	119.662
T_4	13	83.329	14	110.329
T_5	20.333	93.333	17	130.666
T_6	11	57.333	25	93.333
T_7	11.666	57.333	21	89.996
T_8	15.666	40.663	27	83.329
T_9	12.333	17.333	11	40.663
T_{10}	24	17.33	16	57.333
T_{11}	17.333	0	0	17.333

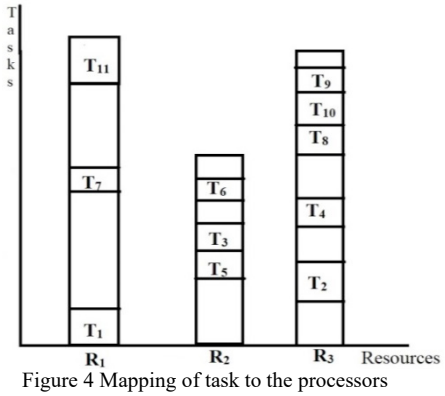

Figure 4 Mapping of task to the processors

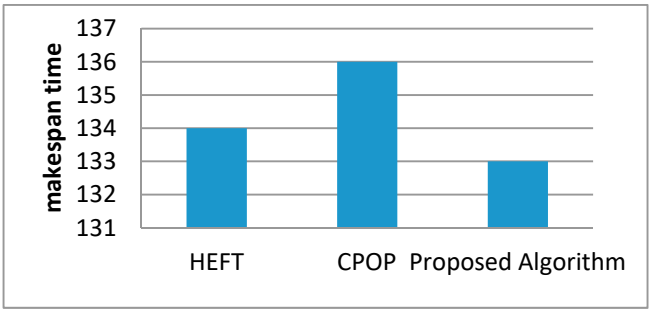

Figure 5 Comparison of makespan time HEFT, CPOP and proposed algorithm

6. Conclusion

We have analyze different scheduling algorithm shown in Table I. Existing task scheduling algorithm works at different and have some pros & cons. In this paper, we modified HEFT algorithm that make a group of tasks based upon their ranking and map the task with processor. After that allocate the tasks to the heterogeneous processor so that it can reduce the makespan time. Evaluated the performance of proposed algorithm at cloudsim simulator, calculated results (Figs 4-5) shows that proposed algorithm reduce the makespan time of tasks and load balancing issues compare to existing HEFT and CPOP algorithm.

References

1. Syed Hamid Hussain Madni, Muhammad Shafie Abd Latiff, Yahaya Coulibaly, Shafi'i Muhammad Abdulhamid, "Resource scheduling for infrastructure as a service (IaaS) in cloud computing: Challenges and opportunities", Journal of Network and Computer Applications, Volume 68, 2016, pp 173-200, ISSN 1084-8045.
2. Abdullahi, M., Ngadi, M. A., Abdulhamid, S. I. M. , "Symbiotic Organism Search optimization based task scheduling in cloud computing environment.", Future Generation Computer Systems, Volume 56, 2016, pp 640-650.
3. Antony Thomas, G. Krishnalal, V.P. Jagathy Raj, "Credit Based Scheduling Algorithm in Cloud Computing Environment", Procedia Computer Science, Volume 46, 2015, pp 913-920, ISSN 1877-0509.
4. S. Raghavan, P. Sarwesh, C. Marimuthu , K. Chandrasekaran, "Bat algorithm for scheduling workflow applications in cloud," International Conference on Electronic Design, Computer Networks & Automated Verification (EDCAV), Shillong, 2015, pp 139-144.
5. B. L. Pan, Y. P. Wang, H. X. Li, J. Qian, "Task Scheduling and Resource Allocation of Cloud Computing Based on QoS", Advanced Materials Research, Vols. 915-916, 2014, pp 1382-1385.
6. G. Wang, H. C. Yu, "Task Scheduling Algorithm Based on Improved Min-Min Algorithm in Cloud Computing Environment", Applied Mechanics and Materials, Vols. 303-306, 2014, pp 2429-2432.

7. Abdulhamid, S.M., Latiff, M.S.A., Idris, “Tasks scheduling technique using league champion ship algorithm for makes pan minimization in IaaS cloud.”, ARPN J. Eng. Appl. Sci., Volume 9, 2015, pp 2528–2533.
8. Bansal, N., Awasthi, A. & Bansal, “Task Scheduling Algorithms with Multiple Factor in Cloud Computing Environment”, Information Systems Design and Intelligent Applications, Springer.
9. Babu, K.R., Samuel, P., “Enhanced bee colony algorithm for efficient load balancing and scheduling in cloud”, Innovations in Bio-Inspired Computing and Applications , 2016, pp 67-78.
10. Shojafar, Mohammad and Javanmardi, Saeed and Abolfazli, Saeid and Cordeschi, Nicola, “FUGE: A Joint Meta-heuristic Approach to Cloud Job Scheduling Algorithm Using Fuzzy Theory and a Genetic Method”, Cluster Computing, Volume 18, pp 829-844.
11. Javanmardi, S., Shojafar, M., Amendola, D., Cordeschi, N., Liu, H. & Abraham, A., “Hybrid Job Scheduling Algorithm for Cloud Computing Environment”, Proceedings of the Fifth International Conference on Innovations in Bio-Inspired Computing and Applications IBICA 2014, 2014. Springer, pp 43-52
12. Chun-Yan Liu, Cheng-Ming Zou, and Pei Wu. “A Task Scheduling Algorithm Based on Genetic Algorithm and Ant Colony Optimization in Cloud Computing”., Distributed Computing and Applications to Business, Engineering and Science (DCABES '14). IEEE Computer Society, Washington, DC, USA, 2014, pp 68-72.
13. H. Yang, "Improved Ant Colony Algorithm Based on PSO and its Application on Cloud Computing Resource Scheduling", Advanced Materials Research, Vols. 989-994, pp. 2192-2195, 2014.
14. G. Le, K. Xu and J. Song, "Dynamic Resource Provisioning and Scheduling with Deadline Constraint in Elastic Cloud," 2013 International Conference on Service Sciences (ICSS), Shenzhen, 2013, pp. 113-117.
K.Dubey et al., “A Priority Based Job Scheduling Algorithm Using IBA and EASY Algorithm for Cloud Metaschedular”, International
15. Conference on Advances in Computer Engineering and Applications, Ghaziabad., India,2015, pp. 66-70.
16. Doaa M., Abdelkader, Fatma Omara, “Dynamic task scheduling algorithm with load balancing for heterogeneous computing system”, Egyptian Informatics Journal, vol. 13, pp. 135-145, july-2012
17. Topcuouglu, Haluk and Hariri, Salim and Wu, Min-you, “Performance-Effective and Low-Complexity Task Scheduling for Heterogeneous Computing”, IEEE Trans. Parallel Distrib. Syst., vol. 13, no. 3, pp. 260-274, Mar. 2002.
18. H. Ren et al., “The load balancing algorithm in cloud computing environment”, International Conference on Computer Science and Network Technology., Changchun., China, 2012, pp. 925-928.”
19. Ergu, Daji, Kou, Gang, Peng, Yi, Shi, Yong, Shi, Yu", "The analytic hierarchy process: task scheduling and resource allocation in cloud computing environment", The Journal of Supercomputing", vol. 64, pp 835--848, jun. 2013.
20. Mala Kalra, Sarbjeet Singh, “A review of metaheuristic scheduling techniques in cloud computing”, Egyptian Informatics Journal, vol. 16,
21. pp. 274-295, aug. 2015.
22. Mohit Kumar, Kalka Dubey, S.C. Sharma, “Job Scheduling Algorithm in Cloud Environment Considering the Priority and Cost of Job”,
23. International Conference on Soft Computing for Problem Solving: SocProS 2016, pp. 313-320.