Intern. J. Fuzzy Mathematical Archive
Vol. 6, No. 2, 2015, 161-168
ISSN: 2320 –3242 (P), 2320 –3250 (online)
Published on 22 January 2015
www.researchmathsci.org

International Journal of
Fuzzy Mathematical
Archive

Analysis of Task Scheduling Algorithm in Fine-Grained and Course-Grained Dags in Cloud Environment

***T.Lucia Agnes Beena*[1] and *D.I.George Amalarethinam*[2]**

[1]Department of Information Technology, St. Joseph's College, Trichy, Tamil Nadu India. e-mail: jerbeena@gmail.com
Corresponding Author

[2]Department of Computer Science, Jamal Mohamed College, Trichy, Tamil Nadu, India
e-mail: di_george@ymail.com

Received 3 November 2014; accepted 30 November 2014

Abstract. Cloud computing environment has the ability to scale-up or scale-down the amount of resources provisioned to an application to balance the demand. This elastic nature of the cloud is also limited by the ownership and locality. The emergence of Hybrid cloud (combination of private and public cloud) arises as a solution. In a cloud computing environment, task scheduling becomes more complex due its transparency and flexibility. As the beneficiaries of cloud are small and medium enterprises, cost of using the resources, forms the important factor in task scheduling. The user applications are represented by the Directed Acyclic Graph (DAG) which includes the communication cost and computation cost of the various tasks in the graph. Some applications are termed as communication intensive (course-grained) and the rest are computation intensive (fine-grained). The aim of this paper is to apply the Customer Facilitated Cost-based Scheduling algorithm (CFCSC) to both communication intensive and computation intensive DAGs. This paper also records the performance of the CFCSC for better service.

Keywords: Cloud computing, task scheduling, DAG, cost-efficient scheduling, CCR

1. Introduction

Cloud computing, a type of parallel and distributed system, delivers the infrastructure, platform and software services as a pay-as-you-go model for the customers. By using clouds, the Information Technology companies are freed from the setting up of basic hardware and software infrastructure. Thus, they can focus on innovation and creation of business values for their application services [1]. The elastic nature of cloud conserves resources as well as money.

In the Cloud, the resource sharing can take beyond the geographical locations. So the Cloud computing environment helps the researcher to share their ideas at the global level. This forms a new way to solve the problems encountered in the scientific workflow. The scientific workflows have very large data set are strongly data dependent. These data sets are to be uploaded in various data centers in different geographical area or in the same data center, so that the tasks of the workflow can be scheduled in parallel. This will lead to the excessive transmission volume and traffic volume [8]. Due to the

frequent transmission, the cost for using the transmission media increases and it also has an impact on the implementation of the scientific workflow. It is also important to study the effectiveness of the task scheduling algorithm in a Cloud environment with respect to the communication cost of the tasks. The Customer Facilitated Cost-based Scheduling algorithm (CFCSC) [7] reduces the computation cost compared to the famous HEFT algorithm. In this paper, it is tested whether the CFCSC algorithm is best suited for communication intensive or computation intensive DAGs.

2. Related work and motivation

The motivation for this paper is from Shishir et al. [13]. They proposed a heuristic approach on the ordering of clean-up jobs of the workflow. A fine and course-grained genetic approach for the Data-intensive workflows was proposed to optimize the schedule. Applying both heuristic and genetic algorithms, they reduced the overall cost as well as the execution time of large Data-intensive workflows for Grid Resources. Arun et al.[2] presented an algorithm for scheduling the workflow tasks to the resources taking into account disk-space constraints and attained a feasible solution for Grid environment. An Algorithm proposed by Chauhan et al. [4] yields schedule based on both the communication cost and computation cost related to tasks. Hence unlike computing field scheduling which is applicable only in case of computation intensive tasks, this new fully decentralized algorithm for Peer to Peer (P2P) grid gives good schedule for tasks; irrespective of the fact whether they are computation intensive or communication intensive in nature. Taura and Andrew [14] designed a heuristic algorithm that maps data-processing tasks onto heterogeneous resources. This algorithm achieves a good throughput of the whole data-processing pipeline, taking both parallelism (load balance) and communication volume (locality) into account. It performs well both under computation intensive and communication intensive conditions. A decentralized scheduling algorithm for Peer to Peer grid systems proposed by Piyush [5] optimizes the schedule compared to the conventional approaches. This algorithm takes accurate scheduling decisions depending on both computation cost and communication cost associated with DAG's subtasks.Cloud computing is an emerging paradigm where traditional resource allocation approaches, inherited from cluster computing and grid computing systems, fail to provide efficient performance. The main reason is that most of cloud applications require availability of communication resources for information exchange between tasks, with databases, or end users. A non-linear programming model to minimize the data retrieval and execution cost of data-intensive workflows in Cloud was formulated by Suraj et al. [12]. In this paper, a comparative study was made between Amazon Cloud Front's 'nearest' single data source selection and the non-linear algorithm. The non-linear algorithm saved three-quarters of total cost. A new communication-aware model for cloud computing applications, called CA-DAG proposed by Kliazovich et al. [9] overcomes shortcomings of existing approaches using communication awareness. Using this model developing a new scheduling algorithm of improved efficiency can be cracked.

3. DAG model

The user application is presented in Directed Acyclic Graph (DAG). DAG is an acyclic graph with nodes and directed edges. Nodes in DAG represent tasks in the application,

and directed edges represent precedence (data dependency) relation between two tasks. A task without any predecessors is called entry node, and the task without any successors are called exit node. The *m-th* predecessor and *n-th* successor of node *i* is denoted as *pre*(*i, n*) and *suc*(*i, n*) , respectively. A DAG normally has only one entry node and exit node. If an application has multiple entry nodes, a node with zero computation time and transmission cost is added to the beginning of DAG as a dummy entry node. In the case of multiple exit nodes, a dummy exit node is inserted in a similar manner.

Formally, a DAG is defined [3] as a tuple $G = (V,E,C, T)$, where V is the set of nodes; E is the set edges e, and eij represents the directed edge from node i to node j; C is the set of computation time, and $C(i)$ denotes the computation time required by executing task i; T is the set of transmission time, and $T(i, j)$ represents the transmission time associated with the edge eij. When node i and node j are scheduled on the same processor, $T(i, j) = 0$.

3.1. Communication to computation ratio (CCR)

The Computation-Communication Ratio (CCR) is another important parameter of DAGs. CCR is defined [10] as the average data transmission time (in cycles) divided by the computation time (also in cycles), as given in Eq. (1).

CCR = Average Data Transmission Time /Average Computation Time (1)

The importance of communication in the task graph and its impact in the scheduling of task to the resources can be better understood with the help of CCR. Based on CCR, the task graphs can be classified as[11], when CCR > 1, the DAG is a communication intensive or coarse grained graph; when the CCR < 1, it is the computation intensive or fine grained graph; when the communication cost and computation cost is equal, i.e., CCR = 1, it is called mixed graph.The literature shows that some conventional algorithms [4] perform better for computational intensive DAGs. In the Cloud scenario, it is better to analyze scheduling parameters of the task scheduling algorithms both in coarse grained and the fine grained graph.

3.2. Random DAGs

A large number of papers on scheduling algorithms use randomly generated DAGs for evaluation. These types of graphs are generated using a number of parameters[11], such as

- the number of tasks in a graph
- communication to computation ratio (CCR)
- the interval from which the costs for communication and processing are randomly selected

This paper uses the random task graphs generated by George et al. [6]. It generates the random tasks along with the communication cost and computation cost for every task in the DAG. The Figure 1 and Figure 2 depict the sample random task graphs generated by the DAGEN tool [6] for computation-intensive and communication-intensive DAGs respectively. The value inside the bubble represents the task number and the value above the directed edges represents the communication cost between the two nodes. The computation cost of the various nodes for communication-intensive and computation-intensive DAGs are tabulated in Table 1.

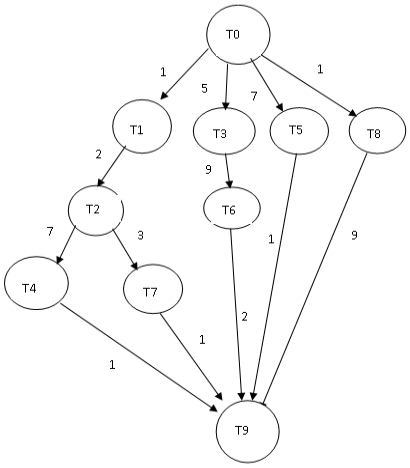

Figure 1: Fine grained DAG

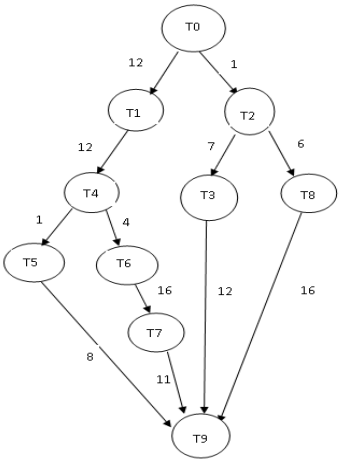

Figure 2: Coarse grained DAG

Communication-intensive DAG		**Computation-intensive DAG**	
TASK	**Computation Cost**	**TASK**	**Computation Cost**
T0	2	T0	8
T1	6	T1	2
T2	4	T2	2
T3	8	T3	11
T4	5	T4	13
T5	2	T5	19
T6	7	T6	10
T7	2	T7	9
T8	3	T8	17
T9	9	T9	10

Table 1: Computation Cost of the DAGs

3.3. Customer facilitated cost-based scheduling algorithm in cloud (CFCSC)

In Cloud computing environment, task scheduling becomes a challenging issue due to its resource virtualization. Hong et al. [8] discuss the specific goals of task scheduling. The goals are load balance, quality of service (QoS), economic principle, the optimal operation time and system throughput. The CFCSC algorithm [7] tries to reduce the computation cost of the customer. It also favours the customer by executing the application at lesser cost compared to HEFT algorithm without affecting or increasing the execution time. In Cloud environment, resources are widely distributed all over the world. Scheduling tasks in various data centre need good communication link for data transfer among the tasks. Hence, this paper inspects the performance of the CFCSC algorithm on the fine grained and coarse grained DAGs. The scheduling parameters

makespan (total execution time) and the cost for execution of the DAG are considered for experiments.

3.2. Results and discussion

The random task graphs needed for the experiments are spawned using the DAGEN tool [6]. The CCR for the fine grained graph is set to 0.4. The number of tasks are varied from 50 to 5000 and correspondingly, the number of resources are also varied from 4 to 60.

CCR < 1					
CCR	**TASKS**	**RESOURCES**	**MAKESPAN**	**COST**	
				HEFT	**CFCSC**
0.44	50	4	101.57	27.65	27.46
0.5	100	4	220.33	32.77	32.14
0.44	150	5	226.62	83.33	82.69
0.44	200	5	297.15	97.32	96.39
0.44	250	6	325.97	129.01	127.3
0.44	300	6	452	79.35	79.16
0.4	350	7	439	120.68	119.16
0.44	400	7	434.33	169.6	168.24
0.4	450	8	483.73	215.37	212.79
0.44	500	8	471.5	235.68	233.58
0.44	550	9	592.6	105.49	105.24
0.4	600	9	537.76	250.08	247.93
0.44	650	10	529.77	276.31	275.17
0.44	700	10	545.97	260.02	259.71
0.4	750	11	548.94	447.94	445.8
0.44	800	11	510.38	476.23	473.9
0.44	850	12	549.73	490.73	489.02
0.44	900	12	531.86	550.61	549.74
0.44	950	13	500.8	587.24	586.81
0.44	1000	13	622.7	367.51	366.82
0.44	2000	26	607.33	707.33	705.23
0.44	3000	39	601.68	1345.11	1344.41
0.44	3500	40	670.95	1787.77	1786.29
0.44	4000	50	653.476	1818.85	1816.13
0.44	5000	60	709.4	2258.87	2257.06

Table 2: Fine grained DAGs (computation-intensive)

The scheduling parameters the makespan and cost are recorded for both the HEFT and CFCSC algorithms. The results are tabulated in Table 2. In the same way, for

coarse grained graph the CCR is taken as 2 and 2.25. The experimental results are tabulated in Table 3.

CCR > 1					
CCR	**TASKS**	**RESOURCES**	**MAKESPAN**	**COST**	
				HEFT	**CFCSC**
2	50	4	89	7.13	6.76
2.25	100	4	157.1	25.49	24.83
2.5	150	5	171.29	28.25	27.34
2	200	5	161.9	70.85	70.57
2.25	250	6	180.57	56.9	56.22
2.25	300	6	212.5	107.02	106.4
2.25	350	7	235.43	87.56	87.01
2.25	400	7	249.83	98.09	97.86
2.25	450	8	231.84	105	104.15
2.25	500	8	273.27	123.06	122.63
2.25	550	9	271.31	111.42	110.49
2.25	600	9	265.69	175.53	174.76
2.25	650	10	295.74	180.49	180.18
2.25	700	10	332.99	163.17	162.49
2.25	750	11	319.69	174.6	174.09
2.25	800	11	320.86	164.42	164.08
2.5	850	12	323.49	158.42	158.08
2.25	900	12	288.26	276.63	276.16
2.25	950	13	346.43	250.46	249.95
2.25	1000	13	326.85	196.59	195.28
2.25	2000	26	354.71	445.66	444.42
2.25	3000	39	345.82	677.15	676.95
2.25	3500	40	410.8	712.25	711.57
2.25	4000	50	367.4	816.89	815.96
2.25	5000	60	405.14	992.07	991.13

Table 3: Coarse grained DAGs (Computation-intensive)

It is observed that the CFCSC algorithm is in line with the HEFT algorithm in makespan irrespective of fine grained or coarse grained DAGs. At the same time, the cost is reduced by 1% in fine grained DAGs. In the coarse grained DAGs, as the number of tasks increase the reduction in cost is reduced to 1% from 5%. The graphical representation of the cost analysis is shown in Figure 3 and Figure 4.

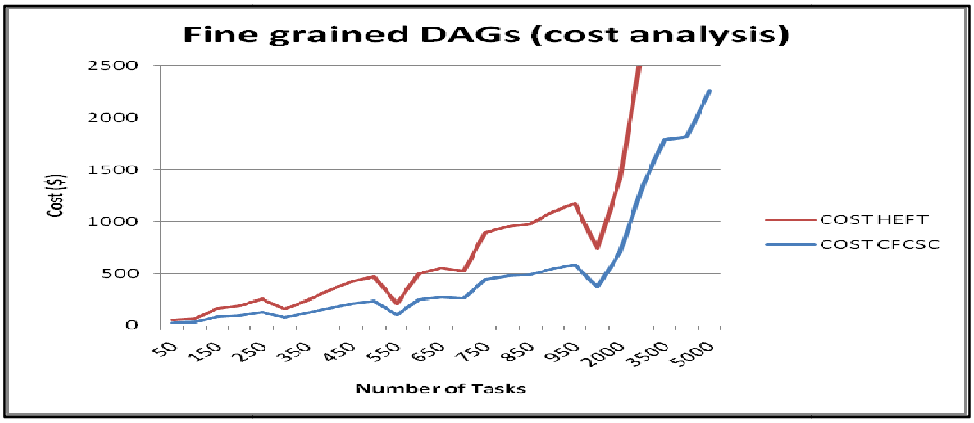

Figure 3: Fine grained DAG

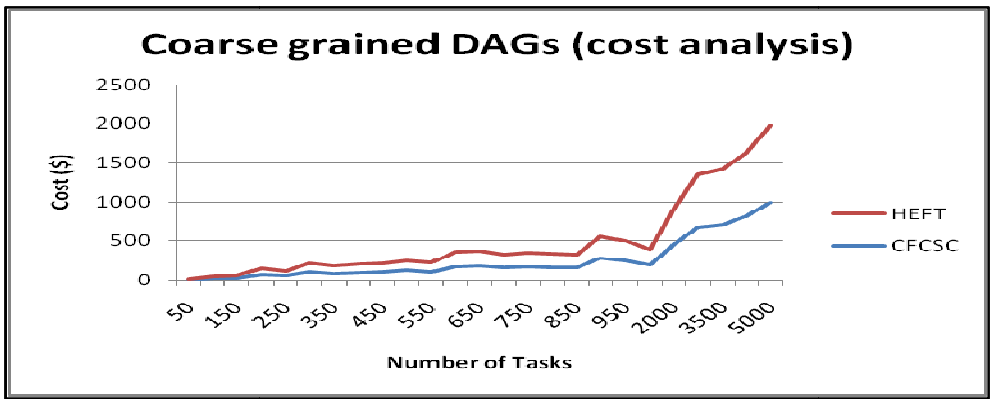

Figure 4: Coarse grained DAG

6. Conclusion

Cloud computing has brought a tremendous change in the business model. The small, medium scale enterprises and public are the customers of Cloud services who take the advantage of the Cloud technology. The famous companies like Google, Amazon, Microsoft Azure, GoGrid, Sun and IBM are competing with each other to provide Cloud services to their customers according to their requirement. This challenges the researchers to develop new ideas for the better service. This paper is one of the initiatives to provide the economic benefit to the customer, while executing their application in the Cloud. The CFCSC algorithm outperforms the HEFT in economic cost without any change in the makespan of the application. This paper applies the CFCSC algorithm to communication-intensive and computation-intensive DAGs. Also it is observed that CFCSC algorithm produces better schedule with lesser cost for fine grained and coarse grained DAGs. In future, bandwidth of the communication links can be included as a scheduling parameter to produce a better makespan.

REFERENCES

1. M.Armbrust, A.Fox, R.Griffith, A.Joseph , R.Katz, A.Konwinski, G.Lee, D.Patterson, A.Rabkin, I.Stoica, M.Zaharia, A view of cloud computing.

Communications of the ACM, 53 (2010) 50–58.
2. A.Ramakrishnan, G.Singh, H.Zhao, E.Deelman, R.Sakellariou, K.Vahi, K.Blackburn, D. Meyers and M.Samidi, Scheduling data-intensiveworkflows onto storage-constrained distributed resources, In Cluster Computing and the Grid, 2007. CCGRID 2007, *Seventh IEEE International Symposium*, 2007, 401-409.
3. C.Song, Y.Li, J.Wang and C.Wu., An Energy-efficient Scheduling Algorithm for Computation-Intensive Tasks on NoC-based MPSoCs., *Journal of Computational Information Systems*, 9 (2013) 1817-1826.
4. P.Chauhan and N. Nitin, Decentralized Computation and Communication Intensive Task Scheduling Algorithm for P2P Grid,InComputerModelling and Simulation (UKSim), 2012 UKSim 14th International Conference, IEEE, 2012, 516-521.
5. P.Chauhan, Decentralized Scheduling Algorithm for DAG Based Tasks on P2P Grid, *Journal of Engineering*, 2014.
6. D.I.George Aamalarethinam and G.J.Joyce Mary, DAGEN – A tool to generate arbitrary directed acyclic graphs used for multiprocessor scheduling, *International Journal of Research and Reviews in Computer Science*, 2 (2011) 782 – 787.
7. D.I.George Aamalarethinam, T.Lucia Agnes Beena, Customer facilitated cost-based scheduling algorithm in cloud, *International Conference on Information and Communication Technologies*, Elsevier Procedia, 2014.
8. H.Sun, S.-ping Chen, C.Jin, K.Guo, Research and simulation of task scheduling algorithmin cloud computing, TELKOMNIKA, *Indonesian Journal of Electrical Engineering*, 11 (2013) 6664 - 6672.
9. D.Kliazovich, J. E. Pecero, A.Tchernykh, Pascal Bouvry, Samee U. Khan, and Albert Y. Zomaya, Ca-dag: Communication-aware directed acyclic graphs for modeling cloud computing applications, In Cloud Computing (CLOUD), 2013 IEEE Sixth International Conference IEEE, 2013, 277-284.
10. Y.-K.Kwok and I. Ahmad, Static scheduling algorithms for allocating directed task graphs to multiprocessors, *ACM Computing Surveys*, 31 (1999) 406-471.
11. A.Olteanu and Andreea Marin, Generation and evaluation of scheduling DAGs: How to provide similar evaluation conditions, *Computer Science Master Research,* 1 (2011) 57-66.
12. S.Pandey, A.Barker, K. K.Gupta and R.Buyya., Minimizing execution costs when using globally distributed cloud services, In Advanced Information Networking and Applications (AINA), 24th IEEE International Conference, 2010, 222-229.
13. ShishirBharathi., Ann Chervenak., Scheduling Data-Intensive Workflows on Storage Constrained Resources, WORKS'09 Proceedings of the 4th Workshop on Workflows in Support of Large-Scale Science, ACM, 2009, 3.
14. K.Taura and A.Chien., A heuristic algorithm for mapping communicating tasks on heterogeneous resources, In Heterogeneous Computing Workshop, 2000.(HCW 2000) Proceedings IEEE, 2000, 102-115.