

Review of Job Scheduling Mechanism in Heterogeneous Grid

Madhu kalotra¹, Richa Dogra²

¹Student, Deptt.of CSE, GCET,Gurdaspur, Punjab , India

²A.P (CSE), GCET,Gurdaspur, Punjab , India

Abstract- Because of a phenomenal increment in the quantity of computing assets in various associations, compelling jobs scheduling algorithms are required for proficient asset use. Job scheduling in considered as NP difficult issue in parallel and disseminated registering situations, for example, group, matrix and mists. Metaheuristics, for example, Genetic Algorithms, Ant Colony Optimization, Artificial Bee Colony, Cuckoo Search, Firefly Algorithm, Bat Algorithm and so on are utilized by researchers to get close ideal answers for work scheduling issues. These metaheuristic algorithms are utilized to plan distinctive sorts of jobs, for example, BSP, Workflow and DAG, Independent undertakings and Bag-of-Tasks. This paper is an endeavor to give exhaustive review of prominent nature-enlivened metaheuristic procedures which are utilized to plan distinctive classifications of jobs to accomplish certain execution targets.

Index Terms- ACO, BAT, Cuckoo, genetic algorithm.

INTRODUCTION

Cloud computing in modern era provides way of using resources without their physical presence at source. The service provided by the cloud is at the front end of computing and internet is at back end. In other words internet is heart and soul of cloud computing. [1]Cloud computing provide mechanism for the users to perform operations that required heavy resources not possessed by them at pay per use basis. With the rapid development of hardware and software cloud computing brings the revolution in the business industry. It provides resources like computational power, storage, computation platform ad applications to user on demand through internet. Some of the cloud providers are Amazon, IBM, Google, Sales force, Microsoft etc. [2]Cloud computing features included resource sharing, multi-

tenancy, remote data storage etc. but it challenges the security system to secure, protect and process the data which is the property of the individual, enterprises and governments. Even though, there is no requirement of knowledge or expertise to control the infrastructure of clouds; it is abstract to the user. It is a service of an Internet with high scalability, quality of service, higher throughput and high computing power. [3]Cloud computing providers deploy common online business applications which are accessed from servers through web browser. Data security is the biggest issue in cloud computing and it is not easy to resolve it. In our review paper we will review the different ways to manage the confidentiality of the data. Before discussing migration mechanisms we discuss services provided by cloud along with types of cloud.

1.1 Cloud Services

[4]There exist legion of services associated with cloud. these services are as described below

• IaaS

[5]Infrastructure as services is critical services provided through cloud. virtualised computing resources are provided by the application of IaaS. Internet is key element with which IaaS is accessed. Cost is encountered on the basis of usage.

• PaaS

[6]Platform as a service is another cloud service that enhance the organizational applications. Large number of applications exists that are supposed to execute over the distinct machines. All the applications has distinct requirements in terms of platform. This platform requirement is accomplished using cloud computing. Cost is encountered on the basis of time period for which platform is online.

• SaaS

[7]Software as a service is another critical service supported through cloud. cloud computing host software which can be accessed by users having access to cloud. in other words machines having limited resources can use SaaS to access software's that they don't possess.

Job Scheduling

Job scheduling (Karatza 2000) is a procedure in which jobs apportioned to the PC are scheduled by utilizing single cluster or multicluster approach. The job allocation process is done in light of constrained resources. In both the approaches we have different algorithms to solve the jobs like FCFS, SJF, Round Robin, and Priority Scheduling in single cluster job. Furthermore, in case of multi cluster job we can use algorithms like ACO, Firefly, Honey Bee, BAT, Cuckoo etc are used to allocate resources so that jobs can be execute in efficient manner and optimal results can be achieves in short span of time.

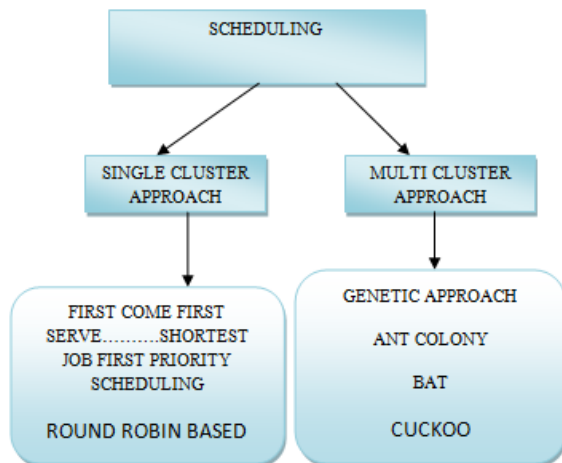


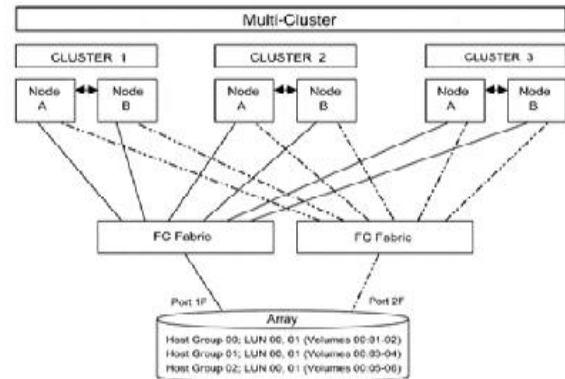
Figure 1: Scheduling in single and Multi Cluster Environment

MULTI CLUSTER ENVIRONMENT

Clusters are gathering of servers with distinct setup. With single cluster, hubs are of comparative arrangements. As (Buyya and Baker 2002)multiple clusters are bound together to form multi cluster condition. In multi cluster condition in this manner, particular nodes(servers) are assembled together. Multi cluster condition is utilized as a part of request to execute complex jobs. These jobs may not be executed by single cluster consequently distribution

methodologies are constructed. Multi cluster condition is appeared as under

Figure 1.2.1: Multi Cluster Environment



Parallel Computing

[8] Parallel computing is working on the rule that enormous issues can be subdivided into littler ones, which are then unraveled in the meantime. In this sort of calculation many jobs are executed at the same time on various processors

The jobs in (Karatza 2000)parallel computing can be characterized by the level at which the equipment support parallelism, with multi-center and multi-processor PCs having numerous handling components inside a solitary machine, while groups, MPPs and lattices utilize various PCs to chip away at a similar job.

Particular parallel PC designs are once in a while utilized close by traditional processors, for quickening particular jobs. Parallel programming models and devices are required for elite computing. The accessibility of multi-center CPUs has given new slant to the common memory parallel programming approach. Parallel computing is operating on the rule that huge problems can frequently be subdivided into smaller ones, which are then solved at the same time. In this type of computation many jobs are executed simultaneously on different processors

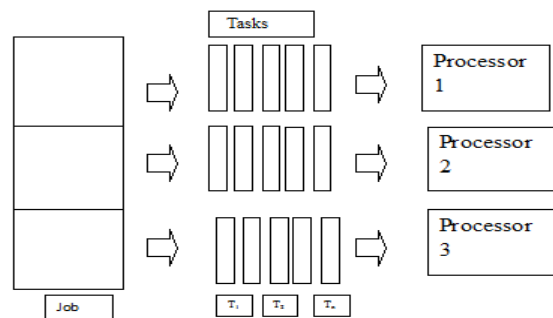


Figure 1.3.1 Parallel Computing Of Jobs

In parallel the jobs are subdivided into tasks and after that those errands are doled out to the different resources as indicated by job necessity. The parallel job calculation is far much superior to the successive calculation in light of the fact that in consecutive calculations the jobs are executed in a solitary line to one processor. So it expends a great deal of time to execute number of jobs .So to determine this issue the jobs are executed in parallel so by doing as such the execution time of jobs are being decreased and expanded the throughput of the framework.

Multiheuristic Approach

Multiheuristic approach is used to NP- hard Problems. These problems are very complex in nature which may not be solved using the single processor system. Their requirements are such that single processor is unable to satisfy the requirements or need of the system. So systems with more than one processors or multi clusters are required. With every job objective is associated. These objectives are required to be satisfied in order to complete the job. Such jobs fall under the category of multiheuristic. The algorithms are as under.

GENETIC ALGORITHM

Genetic algorithm is utilized as a part of request to take care of job schedule issue in multi bunch condition. With each activity target work is related. (Potuzak 2011) Genetic approach continues to fulfill the target work related with the activity. This algorithm ends on the off chance that target work related with each activity is fulfilled or number of ages ends. There are stages related with the genetic algorithm.

Initialization

Jobs and resources at first place required to be initialized. Jobs are known as chromosomes. These chromosomes are initially selected randomly for allocation. Generations, objective function, crossover and mutation probabilities are initialized in this phase.

Selection

In genetic algorithm, Jobs are referred to as Chromosomes. Resources are required by chromosomes. Chromosomes are selected initially for resource allocation. This process is known as selection. In genetic algorithm, this selection is

performed using variety of mechanism known as selection criteria's. Chromosomes are selected for later breeding using this phase. Steps in selection are listed as under

- Fitness function is evaluated associated with each job. Fitness values are normalized by adding the fitness values and then dividing it with total number of jobs.
- Population is sorted according to descending value of fitness values
- Population with highest fitness value is selected for mating.

Crossover or Mating

Selected Chromosomes are then go through this phase in order to determine chromosomes which are to be mutated. Uniform crossover is preferred in proposed thesis.

Mutation

Chromosomes selected for mating in crossover are mutated to generate new chromosomes and then evaluated again.

After performing all the phases, fitness values are analysed again. This process continues until desired level of fitness values is achieved or generations terminates.

ANT COLONY OPTIMIZATION

This is an algorithm which is inspired from the natural behaviour of ants. [9] Ant colony optimization considers the random walk associated with Ants. Ants look for food source and if found laid down pheromone trail to be followed by other Ants. This process is simulated in job scheduling. Ants act as Jobs and Food act as resources. With Ant colony Optimization Convergence is not always guaranteed. In other words optimal solution may or may not be generated. Result in terms of local and global best solution is obtained. Iterations in this algorithm continue until all the Ants gets the food source. Following steps are critical in Ant Colony Optimization

- All the Ants must visit all the resources at least once
- Distant resource is less likely to be visited due to visibility problem.

- In case resource is found by Ant, pheromone is laid down.
- After previous iteration, Pheromone evaporated.
- Process continues until optimal solution is found and lbest is found out or generations terminates.

Convergence problem exists while using Ant colony Optimization.

BAT ALGORITHM

[10] Bat Algorithm was introduced by Yang in year 2010 based on the echolocation behaviour of Bats. To estimate the distance of prey, bat used echolocation to sense the distance. BAT algorithm is totally based on the behaviour of Bats. The bats usually fly randomly with a frequency, loudness, pulse emission, velocity and position to search for their prey. When the bats hunt for the prey, they automatically adjust their parameters like velocity, frequency, loudness and pulse emission based on the distance among them and the prey.

BAT algorithm is used in many fields like in workflow, for independent jobs and for the parallel jobs and has many applications. BAT algorithm is used for scheduling jobs among the number of processors in a given problem to minimize objective function.

By changing the velocities and the position of BATS gives the similarity of another metaheuristic [11] known as the Particle Swarm Optimization algorithm. [10] BAT algorithm is a metaheuristic algorithm used to allocate resources which are detected. It was inspired by echolocation behaviour of BAT. Searching criteria followed is random in nature. It is particularly useful for constrained intensive environment. The balance between exploration and exploitation has to be maintained which can be accomplished through hybrid approach. Pseudo Code for BAT is as under

- Obtain Job Sequence (J1, J2, J3, -----, Jn)
 - Associate Objective Function with Each Job
- $$F(y) = \{y_1, y_2, \dots, y_n\}$$
- Define range of BAT in terms frequency (F) and Loudness (A)
 - Repeat following Steps until $i < \text{Max_Iterations}$
- Generate Solutions by adjusting Existing frequency
Update location of BATS in case solution found
If (Optimal_Solution == True)

Select this solution and replace it with existing solution

End of if

Randomize the BATS flying and relocate the resources.

In case optimal solution is located increase pulse rate and decrease loudness (A).

Rank the BATS and sort the solution.

End of loop

- Output Solution and job Ordering (J)

Cuckoo Algorithm

Cuckoo search is a meta-heuristic algorithm inspired by the bird cuckoo [12], these are the Brood parasites birds. It never builds its own nest and lays their eggs in the nest of other host bird nest. Cuckoo is a best-known brood parasite. Some host birds can employ directly with the intruding cuckoo. If the host bird identifies the eggs that are not their own egg then it will either throw that eggs away from its nest or simply rid its nest and build a new nest.

In a nest, each egg represents a solution and cuckoo egg represents a new and good solution. The obtained solution is a new solution based on the existing one and the modification of some characteristics. In the simplest form each nest has one egg of cuckoo in which each nest have multiple eggs represents a set of solutions. CS is successfully used to solve scheduling problems and used to solve optimization problems in structural engineering. In many applications like speech reorganization, job scheduling, global optimization cuckoo search algorithm is used. Cuckoo search idealized such breeding behaviour and can be applied to various optimization problems

This algorithm is also a part of optimization algorithm which is used in order to produce optimal solution by considering decrease in time and energy consumption in multi objective scenario. The detailed steps are listed as follows:

- Initialize Cuckoo with initial set of eggs.
- Lay eggs in distinct nests
- Some eggs are defected and killed
- If population generated is less than threshold value then check for survival of eggs in nest and obtain profit values.
- Otherwise kill cuckoo in worst area.

- If stop condition satisfied then obtain maximize solution otherwise go to step 2.
- Pseudo code for Cuckoo Algorithm is described as under
- Initialize Jobs(J) for Execution along with number of resources available(R), y=0
- Input number of generations G.
- Define Objective function F(x).
- Define initial population in terms of nest
- Solution is in terms of eggs which must be assigned with rank examined through fitness values
- Repeat while stopping criteria is matched or $y < G$
 $Y = y + 1$

Produce new solution using levy flight of cuckoo
 Calculate Fitness value
 Select the nest randomly
 If ($F_i < F_{i+1}$)
 Replace current solution with the new solution z
 End of if
 Asses fitness and sort the solution with maximum fitness value.
 End of loop
 Produce result in terms of Job Ordering

Table1: comparison of various techniques for job scheduling

ALGORITHM	BASIS	OPERATORS	ADVANTAGES	DISADVANTAGES
Genetic Algorithm (GA)[13]	Based on population size	<ul style="list-style-type: none"> • Crossover • Mutation • Selection • Inversion 	<ul style="list-style-type: none"> • Used for rule extraction • for optimization problems 	<ul style="list-style-type: none"> • Time taken for convergence is more. • No guarantee for global optima.
BAT Algorithm[14]	Based on the echolocation behaviour of virtual bats.	<ul style="list-style-type: none"> • Population size • Loudness parameter • Pulse rate • Maximum number of iterations 	<ul style="list-style-type: none"> • Frequency tuning • Automatic zooming into the region of global solution • Parameter control ensures efficient exploration and exploitation 	<ul style="list-style-type: none"> • Convergence in case of complex problems is slow • Searching in case of complex environment can be misleading.
Cuckoo Algorithm[15]	Based on the	<ul style="list-style-type: none"> • Population size 	<ul style="list-style-type: none"> • Use of levy flights 	<ul style="list-style-type: none"> • actual results rare rather
			factor.	
ACO[16]	Based on natural behaviour of ants	<ul style="list-style-type: none"> • jobs • food • walk 	<ul style="list-style-type: none"> • faster convergence as compare to GA 	<ul style="list-style-type: none"> • finding local and global conclusion can be complex

Table1:comparison of various techniques for job scheduling

Cuckoo Algorithm[15]	Based on the property of brood parasitism of cuckoo species	<ul style="list-style-type: none"> • Population size • Switching probability • Step-size scaling factor • Levi exponent 	<ul style="list-style-type: none"> • Use of levy flights result in efficient exploration of search space • Ensures the property of global convergence due to the use of switching probability 	<ul style="list-style-type: none"> • actual result is rare rather approximate solution is produced
----------------------	---	---	---	---

Make Span- Make span is the time to complete the job it includes the starting and finishing time of job. In the whole time job accomplish its task that is to be executed to achieve optimal results.

$$Makespan = \sum finished_i / Total_jobs$$

Flow Time- Flow time is the average amount of time to taken by job to achieve the optimal result. Every single unit in job can be executed within that time.

Waiting Time – The amount of time for which job have to wait for the processor to accomplish its task.
 $waiting_time = Starting_time - Arrival_time$

Execution Time- Execution time is the total amount of time to execute single job.

$$execution_time = finish_time - start_time$$

Normalization Function- Normalization function is used in order to optimize the result which is obtained through metaheuristic algorithm.

Fitness Function- This function is used in order to analyse the fitness of the population or chromosomes. The population will alter according to this fitness function.

CONCLUSION

This paper gives a wide survey of various sorts of metaheuristic strategies which are utilized by scientist to get close ideal arrangements of four unique sorts of jobs in particular BSP, Workflow and DAG, Independent assignments and Bag-of-Tasks. The different issues and issues found in the individual metaheuristics can be overwhelmed by utilizing the hybridization of at least two

metaheuristic procedures. In this paper, an endeavor has been made to give the review and comparison of various metaheuristic algorithms which are utilized to schedule singular kind of jobs.

REFERENCES

- [1] S. Asif, R. Shah, A. H. Jaikar, and S. Noh, "A Performance Analysis of Precopy, Postcopy and Hybrid Live VM Migration Algorithms in Scientific Cloud Computing Environment," pp. 229–236, 2015.
- [2] H. Wang, Z. Kang, and L. Wang, "Performance-Aware Cloud Resource Allocation via Fitness-Enabled Auction," *IEEE Trans. Parallel Distrib. Syst.*, vol. 27, no. 4, pp. 1160–1173, Apr. 2016.
- [3] T. Chalermarwong, T. Achalakul, and S. C. W. See, "The Design of a Fault Management Framework for Cloud," 2012 9th Int. Conf. Electr. Eng. Comput. Telecommun. Inf. Technol., pp. 1–4, 2012.
- [4] I. P. Ekwutuoha, S. Cheny, D. Levy, B. Selic, and R. Calvo, "Energy efficient fault tolerance for high performance computing (HPC) in the cloud," *IEEE Int. Conf. Cloud Comput. CLOUD*, pp. 762–769, 2013.
- [5] J. Guitart, M. Macias, K. Djemame, T. Kirkham, M. Jiang, and D. Armstrong, "Risk-driven proactive fault-tolerant operation of IaaS providers," *Proc. Int. Conf. Cloud Comput. Technol. Sci. CloudCom*, vol. 1, pp. 427–432, 2013.
- [6] C. Pahl and I. Centre, "Containerization and the PaaS Cloud," 2015.
- [7] F. Doelitzscher, A. Sulistio, C. Reich, H. Kuijs, and D. Wolf, "Private cloud for collaboration and e-Learning services : from IaaS to SaaS," pp. 23–42, 2011.
- [8] A. Kaur, "A Review on Various Job Scheduling Algorithms," vol. 13, no. 3, pp. 359–367, 2017.
- [9] S. Lorpunmanee, M. Sap, A. Abdullah, and C. Chompoo-inwai, "An ant colony optimization for dynamic job scheduling in grid environment," *Int. J. Comput. Inf. Sci. Eng.*, vol. 1, no. 4, pp. 207–214, 2007.
- [10] A. Hossein and G. X. Yang, "Bat algorithm for constrained optimization tasks," pp. 1239–1255, 2013.
- [11] M. Widmer, A. Hertz, and D. Costa, "Metaheuristics and Scheduling," *Prod. Sched.*, pp. 33–68, 2010.
- [12] I. Engineering, "OPTIMAL POWER FLOW USING CUCKOO OPTIMIZATION ALGORITHM," pp. 4213–4218, 2013.
- [13] S. Tahilyani, "A New Genetic Algorithm Based Lane-By-Pass Approach for Smooth Traffic Flow on Road Networks," vol. 1, no. 3, pp. 32–36, 2012.
- [14] B. Sotomayor, K. Keahey, and I. Foster, "Combining Batch Execution and Leasing Using Virtual Machines," pp. 87–96.
- [15] N. Optimisation, "A comprehensive review of cuckoo search : variants and hybrids Iztok Fister Jr .*, Dušan Fister and Iztok Fister," vol. 4, no. 4, 2013.
- [16] W. Wen, C. Wang, D. Wu, and Y. Xie, "An ACO-Based Scheduling Strategy on Load Balancing in Cloud Computing Environment," 2015.