ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

A Survey on Modified GA Based Load Balancing Based on Unreliable Resources in Cloud Computing Environment

Khushbu Agarwal[1], Bhopender Kumar[2]

[1]Research Scholar, Dept. of Computer Science and Engineering, JIIT, Noida, India

[2]Research Scholar, Dept. of Computer Science and Engineering, NITTTR, Punjab University, Chandigarh, India

ABSTRACT: Now-a-days cloud computing is the most raising technology due to its elasticity of resource provisioning and the pay-as-you-go pricing model which enables users to pay only according to their need. The cloud can be accessed anytime and anywhere. Cloud computing is based on virtualization technology for providing various virtual machines over a physical hardware. That increase availability of resources. So it should provide high performance to the user and good benefit to the cloud service provider. So there are many challenges in cloud computing. Load balancing is an important one of them. That maximizes the optimum use of cloud resources. To balance the load in cloud systems, the resources and workload must be scheduled in a good fashion so that maximum use of available resources can be made. There are various scheduling algorithms used by the load balancer to schedule the tasks that are running on any virtual machine. Task scheduling helps in the increase optimal resource utilization and avoid under utilization and over utilization of resources. In this paper, we discussed various task scheduling algorithms to resolve scheduling. It brings out an exhaustive survey of such strategies in cloud computing and includes a detailed classification of them.

KEYWORDS: Cloud computing, Workflow scheduling algorithms, Load Balancing, Genetic Algorithm.

I. INTRODUCTION

A. Cloud Computing

Today is the era of smart computing. Everybody wants to use resources instantly. So researchers started to think about a technology, which can serve anywhere anytime. Cloud computing is the latest technology that provides computational resources, storage and many more computing services on a pay per basics. Cloud computing provides all services on the bases of virtualization in which cloud provider provides virtual machine to the user on his demand[1].

B. Workflow Scheduling and Load Balancing

In workflow scheduling a big task is divided in sub tasks and resources are allocated to each of the subtask to be executed successfully in a well scheduled manner. Resources are allocated in such way to achieve some pre-defined objective. Every workflow has a parent child relationship, in which each task is treated as a node and every edge represents the relationship between each node. The parent task is linked to a child task according to set of rules [2]. A workflow application is represented in the form of a direct acyclic graph (DAG) such as G (V, E). Where V represents the number of task and E represent the connection between these tasks. A task with no parent is known as entry task and a task with no child is known as exit task. The workflow scheduling algorithms can be heuristic or meta-heuristic in nature as described in [3]. Workflow scheduling allocates resources or virtual machines to different tasks.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

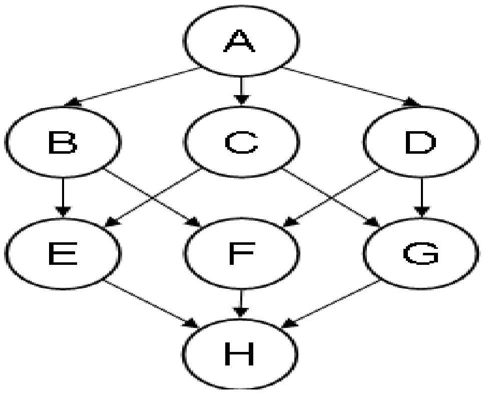

Figure 1: DAG of Eight Dependent Tasks

Load balancing is the process of distributing the load among assorted resources on any system such as a computer cluster, CPUs, etc. Thus load need to be distributed over the resources in a cloud-based architecture, so that each resource does roughly the equal amount of task at any point of time. It is the process of shuffling task from overloaded virtual machine to under loaded virtual machine in the cloud system to make utilization effective and improve the response time of the task. Any load balancing algorithm will be aiming to optimize resource usage and maximize the response time and avoid overload of any virtual machine. Thus the load balancing problem is complex and can be considered as a computationally inflexible problem. Such a problem cannot be formulated by linear programming hence it is quite difficult to find the globally optimal solution by using deterministic polynomial time algorithms or rules.

Based on the task dependency, the tasks can be classified as independent and dependent tasks. The tasks which do not need any communication between the tasks are called independent tasks. The dependent tasks differ from the independent tasks as the previous have a precedence order to be followed during the scheduling process. The main objective in scheduling the dependent tasks is to minimize the make span which is the total length of the schedule, by decreasing the time taken to execute each node called the Computation cost and the communication cost, which the time is taken to transfer data between the two nodes. Thus, the task dependency plays an essential role in deciding the appropriate scheduling strategy.

The rest of the paper is organized as follows. Section 2 explains related work in this field. Section 3 explains proposed work and methodology. Finally, Section 4 concludes this paper with future work.

II. RELATED WORK

Numbers of authors have done work in the area of workflow scheduling and load balancing. Table 1 represents the description of work done in terms of type of scheduling algorithm, nature of scheduling algorithm, factor considered and tools used to simulate the scheduling algorithms were applied. There has been little research on load balancing techniques in a cloud computing environment. The look ahead, genetic algorithm (LAGA) [4] uses genetic algorithms to reduce makespan and consider reliability also. A Novel Heuristic Genetic Load Balancing Algorithm in Grid Computing [5] perform load balancing and reduce make span in grid environment. A Modified Genetic Algorithm for Load Balancing in Heterogeneous Distributed Computing Systems [6] perform load balancing using modified genetic algorithm in Heterogeneous Distributed Computing Systems and reduce make span as a result. A Genetic Algorithm (GA) based Load Balancing Strategy for Cloud Computing [7] perform load balancing and reduce make span in a cloud environment using a simple genetic algorithm.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

Qos Constraints / Scheduling Algorithms	Make Span	Cost	Budget Constrained	Deadline Constrained	Reliability	Energy Efficiency	Scheduling success rate	Resource Utilization	Load Balancing
Hybrid[11]	√	√	X	√	X	X	√	X	√
Modified Genetic[6]	√	√	√	√	√	X	X	√	X
List Scheduling[12]	√	√	√	√	√	√	X	X	√
Particle Swarm Optimization [13]	√	√	√	√	X	√	X	√	√
Intelligent Water Drops[14]	√	X	X	X	X	X	X	X	X
Partial Critical Path[15]	X	√	X	√	X	X	X	X	X
Ant Colony Optimization [16]	√	√	√	√	X	X	X	√	√

Table 1: Table Showing Different Areas Which Require Further Attention and the Areas Which Have Already Been Explored

III. Proposed Work

A. Simple Genetic Algorithm

The algorithm starts by generating an initial population of random candidate solutions. For load balancing problem, each individual in the population represents a random job of the application tasks onto the processors of the distributed system. Each individual is then awarded a score based on its performance. The individuals with the best scores are chosen to be parents. The parents are cut and spliced together using crossover to make children. The generated children are mutated based on a mutation rate, then scored, and the best individuals are chosen to be the parents of the next generation. At some point the process is terminated and the best scored individual in the population is taken as the final result.

Typically, a genetic algorithm has the following steps.

Step 1: Creation of individuals in initial population which is generally done randomly. The size of the initial population depends on the nature of the problem.

Step 2: Generation of new individuals by applying selection method. Rank based roulette wheel selection method will be used which gives better results for larger size problem [8]. In this, parent individuals are selected and crossover and mutation operations are applied to generate new off springs.

Step 3: Evaluation of the fitness value of newly generated individual. The following formula shows the fitness function for the individual.

Task execution time (Tet) = (Instruction length of task) / (MIPS rate of virtual machine)

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

Makespan of individual M (I) $=\sum_{i=0}^{n}$ Teti …..(1)

Fitness function (F) = (Max time) / M (I) …..(2)

Repetition of the step 2 and 3 until stopping criteria is met and then best individual in returned.

B. Proposed methodology

In initial population, individuals are generated randomly. The reliability of the randomly selected virtual machine is checked before scheduling tasks on it. The task t is selected from set T that has not been allocated to any virtual machine and select the VM randomly form set VM. If selected VM is reliable then schedule the task to on VM otherwise select another VM randomly from set VM. This process continues until the individual is generated. The numbers of individuals are generated in the initial population, according to the size of the initial population. The quality of individuals obtained in the initial population depends upon the threshold. The threshold value is set according to workflow application requirement. If the workflow application could not be afforded to fail, then the value of reliability threshold should be higher. This selects only those virtual machines for scheduling which are highly reliable. In evaluation step, the fitness function is calculated for those individuals whose cost is less than workflow application budget. The fitness function is calculated on the basis of makespan [9]. The fitness value is higher, for that individual which has minimum (Equation 1) makespan than other individuals. The workflow application consists of interdependent tasks. The execution time of a task depends on virtual machine MIPS (Million Instructions Per Second) rate. The MIPS rate of a virtual machine represents the computational performance or the number of instructions a virtual machine can execute in one second. The VM that has a high MIPS rate need less time to execute the tasks that are scheduled on it. The M (I) is the makespan of individual which shows the total execution time of all tasks in a workflow application. The Max time is the maximum completion time of individual of the current population. The fitness function (F) for new individual is calculated with the help of maximum completion time for individual in current population and the makespan of a new individual (Equation 2). In the selection process, two individual are selected from the population through rank based roulette wheel selection method. Linear search is performed when individuals are selected. Crossover of two individuals will result in a new individual. Therefore, we need to twist the wheel twice to pick two individuals randomly for crossover. With each twist, the individual under the wheel's marker is selected to act as parent for the next generation. In crossover process, genes of two individuals are interchanged to produce a new individual. The proposed algorithm uses single point crossover where two parent individuals transfer their genes at corresponding point and produce a new individual by selecting the first half from first parent and the second half from second parent. In mutation process, random change is performed in child schedule. Two randomly picked tasks interchange their virtual machines in child schedule. The algorithm terminates when number of generations reach user given threshold. The number of iterations to find the best schedule depends upon threshold value. The genetic algorithm returns better quality individual (schedule) if the threshold value is set higher. After selecting the best schedule we perform load balancing if required in case of high execution rate and the failure of any virtual machine.

Proposed methodology would consider the following:

1. Rank based roulette wheel selection method for population selection.
2. Complexity Minimization for large size real workflow problems like Montage.
3. Balance the load in two situations
   i) when virtual machines are overloaded
   ii) failure of any virtual machine.

GA [10] is considered as one of the most widely used artificial intelligent techniques used mainly for effective search and optimization. It is a stochastic searching algorithm based on the mechanisms of natural selection and genetics. GAs has been proven to be very efficient and stable in searching out global optimum solutions, especially in complex and/ or vast search space.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

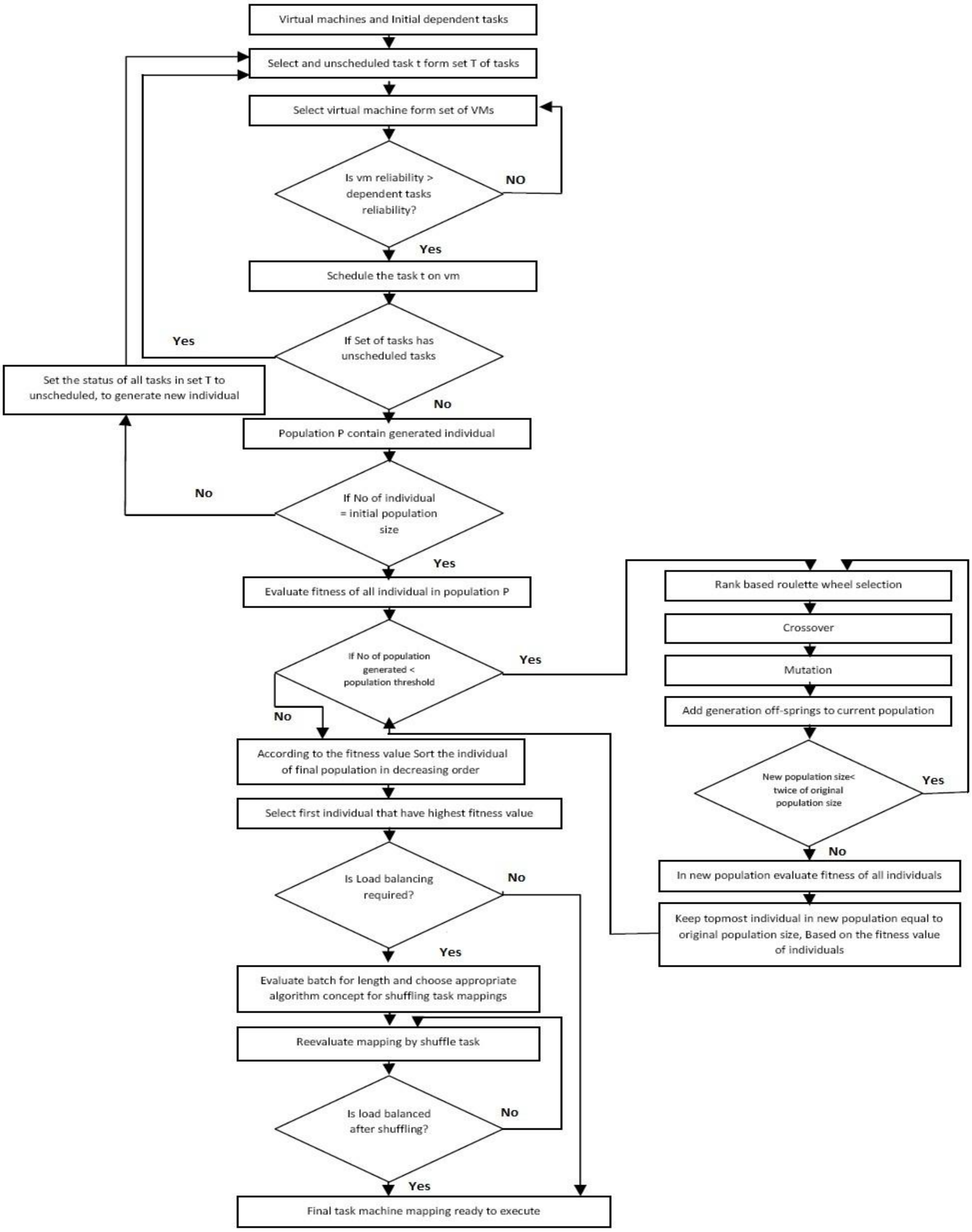

Figure 3: Proposed Methodology

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

IV. Conclusion and Future Work

In this paper, we review and surveyed many workflow scheduling algorithms and load balancing techniques and result is concluded in the form of Table1 on the basis of some factors like algorithm name, nature of the algorithm, factor considered by the research and tools used to implement the research work. Table 1 shows the various research issues in the field of workflow scheduling and load balancing techniques. We have given the overview of heuristics based methods for scheduling like PSO, ACO and Genetic Algorithms etc. From the literature review, we can say there has been done lots of work in this field but still there are many research issues to be considered. We also propose a modified genetic algorithm based methodology for reduce overall execution time and perform load balancing if required. The proposed methodology is depicted by a flow chart also. One of the major issues of cloud computing is load balancing because overloading of a system may lead to poor performance which can make the technology unsuccessful. So there is forever a requirement of efficient load balancing algorithms for efficient utilization of resources.

In our future work we will work on proposed methodology and will try to reduce overall execution time and will perform load balancing while we are considering unexpected makespan or any virtual machine failure.

References

1. Quarati A., Clematis A., Galizia A., Agostino D. D., Mangini M.,'Delivering cloud services with QoS requirements: an opportunity for ICT SMEs', 9th International Conference on Economics of Grids, Clouds, Systems, and Services (Springer, Berlin, 2012), Page No. 197–21
2. Allen R.,'Workflow: an Introduction Workflow Management Coalition, Workflow Handbook,' (2001).
3. Yu J., Ramamohanarao K., Buyya R. K.,'Workflow Scheduling Algorithm for Grid Computing, Meta-heuristics for Scheduling in Distributed Computing Environment', Vol. 146,(Springer Berlin Heidelberg, 2008),Page No. 173-214.
4. Yeo C. S., Su J., Buyya R., Wang X.,'Optimizing the Makespan and Reliability for Workflow Applications with Reputation and a Look ahead Genetic Algorithm', Journal Future Generation Computer Systems, vol. 27, issue 8,(2011), Page No. 1124-1134.
5. Ma J.,'A Novel Heuristic Genetic Load Balancing Algorithm in Grid Computing', Second International Conference on Intelligent Human-Machine Systems and Cybernetics(2010),Page No. 166 - 169.
6. Attiya G. M., Shouman M., Morsi Z.,'A Modified Gen etic Algorithm for Load Balancing in Heterogeneous Distributed Computing Systems', Minufiya J. of Electronic Engineering Research (MJEER),(January 2011), Vol. 21, No.1.
7. Mandal B., Mondal J. K., Dasgupta K., Dutta P., Dam S.,'A Genetic Algorithm (GA) based Load Balancing Strategy for Cloud Computing', International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA) (2013), Pages No. 340–347.
8. Geraghty J., Razali N. M.,'Genetic Algorithm Performance with Different Selection Strategies in Solving TSP', Proceedings of the World Congress on Engineering ,(London, U.K 2011), Vol II WCE, Page No. 5-9.
9. Singh L., Singh S.,'A Genetic Algorithm for Scheduling Workflow Applications in Unreliable Cloud Environment', Recent Trends in Computer Networks and Distributed Systems Security Communications in Computer and Information Science Volume 420,(2014), Page No. 139-150.
10. Goldberg D. E.,'Genetic algorithms in search, optimization, and machine learning', Addison-Wesley, (1989),ISBN: 0201157675.
11. Wang H., Cui L., Xu M., Bi Y.,'A multiple QoS constrained scheduling strategy of multiple workflows for cloud computing', IEEE international symposium on parallel and distributed processing with applications, (2009), Page No. 629-634.
12. Fard H. M.,Barrionuevo J. J. D, Prodan R, Fahringer T.,'A Multi-Objective Approach for Workflow Scheduling in Heterogeneous Environment, Cluster', Cloud and Grid Computing 12th IEEE International Conference(2012), Page No. 300-309.
13. Wu L., Buyya R., Pandy S., Guru S. M.,'A Particle Swarm Optimization-Based Heuristic for Scheduling Workflow Application in Cloud Computing Environments', Advance Information Networking and Applications: IEEE International Conference(April 2010), Page No. 400-407.
14. Nee A. Y. C., Niu S. H., Ong S. K.,'An Improved Intelligent Water Drops Algorithm for Achieving Optimal Job Shop Scheduling Solution', International Journal of Production Research, vol. 50, (Jun. 2012), Page No. 4192–4205.
15. Pema D. H. J. E, Naghibzadeh M., Abrishami S.,'Deadline-Constrained Workflow Scheduling Algorithm for Infrastructure as a Service', Future Generation Computer Systems, Vol. 29, Issue 1,(2013), Page No. 158-169.
16. Yildirim M. B., Barut M., Keskinturk T.,'An Ant Colony Optimization Algorithm for Load Balancing in Parallel Machines with Sequence-Dependent Setup Times', Computer and Operations Research, vol. 39, (2012), Page No. 1225-1235.