Volume 8, No. 5, May-June 2017

ISSN No. 0976-5697

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Research Paper on Genetic Based Workflow Scheduling Algorithm in Cloud Computing

Er. Jyoti Garg
Research Scholar
Dept. of Computer Eng.
Punjabi University
Patiala, India

Er. Gurjit Bhathal
Assistant Professor
Dept. of Computer Eng.
Punjabi University
Patiala, India

Abstract: In the current scenario, cloud computing has emerged widely in companies and enterprises. Cloud enables the organization to utilize hardware, software and applications (e.g. CPU, Memory, Bandwidth, Network, and Storage) based on virtualization and pay-as-per-use principle. But the main challenge before cloud is resource management that is how efficiently and effectively the computer resources should be managed so that no computing device is underutilized or over utilized in the dynamic environment. In this paper, a task scheduling algorithm has been introduced which allocates and executes user's tasks. This proposed algorithm is based on genetic algorithm (GA). The main goal of this algorithm is to reduce completion time and execution cost of different tasks and maximize resource utilization. CloudSim Toolkit is used to evaluate the performance of the proposed algorithm.

Keywords: Cloud Computing; Workflow scheduling; Genetic Algorithm

I. INTRODUCTION

Cloud computing is a new developing paradigm of distributed computing. Cloud is essentially a bunch of commodity computers that are networked together in same or different geographical locations, operating together in order to serve number of customers according to their demands with the help of virtualization [11]. Thus cloud consumers use the services provided by the cloud providers to built their applications in internet and deliver the same to the end users based on the service level agreements (SLA) between the service providers and service consumers.

Cloud computing faces number of challenges like security, performance issues, resource management, reliability etc. But workflow scheduling is one of the main issue in the management of workflows in cloud computing. Scheduling is a process that allocates interdependent tasks on dispersed resources and manages their execution [1]. Workflow scheduling is responsible for allocating suitable resources to workflow tasks. Execution of these tasks can be completed to satisfy quality of service specified by users.

A workflow scheduling algorithm is very necessary to achieve the main goal of cloud computing that is to satisfy cloud users with the quality of service which is agreed in SLA [10]. There are many parameters which must be considered to develop a good workflow scheduling algorithm. Some of these parameters are important from the cloud user point of view like task completion time, execution cost, response time. Other parameters are important from cloud provider point of view like resource utilization, fault tolerant and power consumption. The main objective of this paper is to build a task scheduling algorithm which is based on Genetic Algorithm which allocates and executes tasks in dynamic environment to improve task completion time, reduce execution time and at the same time to increase resource utilization.

Rest of the paper is organized as follows: Section 2 describes the related work. Section 3 discusses the default GA algorithm and proposed GA based workflow scheduling algorithm. Section 4 includes configuration of the CloudSim Simulator and implementation and results of proposed Genetic Algorithm. In the last section, conclusion and future work are given.

II. RELATED WORK

In current years, the hassle of workflow scheduling on a dispensed surroundings has caught the attention of researchers. The main problem is the execution time which has to be minimized. However, scheduling of tasks is taken into consideration a vital problem within the Cloud computing surroundings by using different factors like task completion time, the total execution cost for all tasks, resource utilization, power consumption, and fault tolerance. Lot of work has been done in this area. This section discusses some relevant work done in this field.

R.P. Brent [2] has proposed a strategy known as first fit which is used by many cloud computing systems such as Eucalyptus. This strategy solves the starvation problem but makespan is reduced for the tasks. Also, this method does not support the optimal usage of the resources as requests will execute on each resource.

GE Junwei [5] has supplied a static genetic set of rules by way of thinking about general task completion time, average task completion time, and cost constraint.

In 2010, Yujia Ge and Guiyi Wei [6] proposed a new scheduler which makes the scheduling decision through evaluating the entire institution of tasks in a job queue. A genetic algorithm is designed as the optimization method for a brand new scheduler who affords better makespan and better balanced load throughout all nodes than FIFO and delay scheduling.

In [9], S. Singh has furnished a tricky idea about GA with the aid of introducing numerous versions for project scheduling inside the Cloud computing environment. He has brought a set of rules to solve undertaking scheduling hassle with the aid of modifying GA wherein preliminary population is generated via Max-Min approach to get more optimum consequences in terms of "makespan".

P.kumar [12] proposed an improved genetic algorithm as task scheduling algorithm that is used for efficient utilization of the computing resources in the system.

One of the scheduling troubles is to allocate the best resource to the arrival tasks. The dynamic scheduling procedure is considered complicated if numerous tasks arrive at the same time, therefore, S. Ravichandran and D. E. Naganathan [15] have introduced a model to keep away from this trouble with the aid of allowing the arrived tasks to wait in a queue and the scheduling will recomputed and sort those tasks. Consequently, the scheduling is executed with the aid of taking the first task from the queue and allocated to the available resource on the way to be the nice in shape using GA. The objective of this system is to maximise utilization of resources as also lessen execution time.

In 2011, S. Sindhu & Saswati Mukherjee [16] proposed algorithms for cloud computing surroundings and in comparison it with default policy of cloudsim toolkit while considering computational complexity of jobs.

Furthermore, one of the important dreams of task scheduling from the angle of a Cloud company is to maximize the income by way of making use of resource efficiently. Therefore K. Thyagaarajan, et al. [17] have delivered a version for assignment scheduling within the Cloud computing environment for an powerful advantage of profit at the Cloud computing provider company.

A comparative look at of 3 venture scheduling algorithms at the Cloud computing environment -round-robin, pre-emptive priority and shortest remaining time first algorithms - has been achieved in [18].

Z. Zheng, et al. [19] have proposed an set of rules based totally on GA to address scheduling trouble in the Cloud computing environment known as Parallel Genetic algorithm (PGA) to achieve the optimization or sub-optimization for Cloud scheduling issues mathematically.

III. THE PROPOSED GENETIC BASED WORKFLOW SCHEDULING ALGORITHM

The Cloud provider must assure optimal scheduling of cloud user's tasks inside the Cloud computing surroundings in line with SLA. On the same time, he ought to assure the improved throughput and accurate usage of the Cloud resources. Consequently, the Cloud provider wishes an amazing algorithm to agenda the customers' tasks on the Cloud to fulfil QoS, decrease makespan, and guarantee suitable utilization of the Cloud resources [10]. Therefore, task scheduling is classed as an optimization problem. It is well known fact that workflow scheduling is a NP-complete problem. So, the optimization problems can be solved using heuristic algorithm such as Genetic Algorithm (GA), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO).

In this paper, the proposed workflow scheduling algorithm in the Cloud surroundings is based totally on the default GA with a few adjustments. In keeping with these modifications, the parents can be taken into consideration in every population beside the produced child after the crossover process. Also, the tournament selection is used to pick the exceptional chromosomes to triumph over the trouble of the population length. Tournament selection is a method of choosing an individual from a population of individuals in a genetic set of rules. Tournament selection involves running numerous "tournaments" among a few people (or 'chromosomes') selected at random from the population. The winner of each match (the one with the fine fitness) is selected for crossover. Therefore, the proposed algorithm is called Tournament Selection Genetic Algorithm (TS-GA).

A. Genetic Algorithm

GA was first introduced by Holland in 1975 [13] and represents a population based optimization method based on a metaphor of the evolution process observed in nature. In GA, each chromosome (individual in the population) represents a possible solution to a problem and is composed of a string of genes. The initial population is taken randomly to serve as the starting point for the algorithm [11]. A fitness function is defined to check the suitability of the chromosome for the environment. On the basis of fitness value, chromosomes are selected and crossover and mutation operations are performed on them to produce offspring's for the new population. Principals involved in Genetic algorithm are as follows:

1) *Initial Population:* The initial population is the set of all individuals which may be string of bits (0 or 1) or real numbers that are used inside the GA to find out the highest quality solution. Each solution in the population is called as an individual. Every individual is represented as a chromosome for making it suitable for the genetic operations. From the initial population, the individuals are decided on, and some operations are implemented on them to shape the next generation. The mating chromosomes are decided on primarily based on a few precise criteria.
2) *Fitness Function:* The fitness function generates a fitness value of each chromosome. The value indicates whether a chromosome is suitable for solving a task scheduling problem or not. An individual productivity depends on fitness value. The fitness function is based on execution time of tasks for general task scheduling problems. The performance of an individual is shown by the fitness value in population. Therefore, the motivating factor in GA is fitness function.
3) *Selection:* The selection mechanism is used to pick out an intermediate solution for the next generation based totally at the Darwin's regulation of survival. This operation is the guiding channel for the GA based totally at the performance. There are various selection strategies to pick out the excellent chromosomes which include roulette wheel, Boltzmann method, tournament selection, and selection primarily based on rank.
4) *Crossover:* The crossover operation is a simple method to swap one part of a chromosome for that of other chromosome. This operation can be achieved by choosing two parent individuals and then generating new population on the basis of fitness value. Crossover involves the exchange of the genes in chromosomes that are allowed to reproduce. Crossover can be uniform, single point, or two point crossover.

5) *Mutation:* Mutation takes place after the crossover operation. Mutation helps in generation of new population. The main aim of mutation operation is to expand the search space by changing one part of a chromosome. This operation takes place when the population tends to become homogeneous due to repeated use of reproduction and crossover operators. This principal can produce the entire new gene values which are being added to the gene pool. The genetic algorithm may be able to produce a better solution than was previously because of these newly generated values. Generally, mutation probability keeps very low.

Figure 1 describes the process of genetic algorithm diagrammatically:

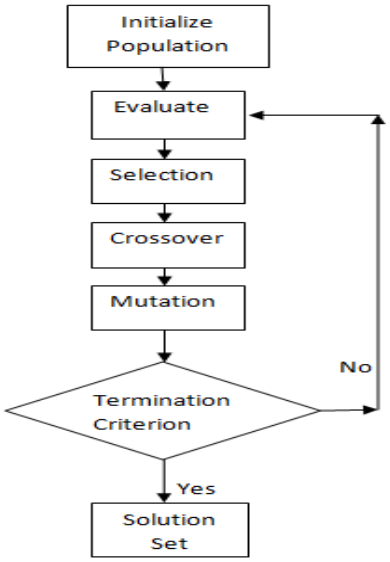

Figure 1. Genetic Algorithm [14]

B. *The Proposed Tournament Selection Genetic Algorithm (TS-GA)*

In this work, a modified GA is proposed to clear up task scheduling hassle in Cloud computing surroundings to enhance completion time for executing all jobs on the VMs, inside the same time, decrease the entire cost of utilization the resource and maximize utilization of the cloud resources [8]. The principle concept of this proposed algorithm (i.e., TS-GA) is that once every choice inside the population, there's a solution that could fulfil good fitness feature, however it is not selected to crossover technique. Through the proposed algorithm, this solution is not removed from the population, but it is selected and introduced to the population while subsequent new release is started. This step is considered as an amazing step as a number of the iterations can generate the optimized solution.

1) *Initialize Population:* The population is randomly generated using encoded binary (0, 1) according to the proposed tournament selection genetic algorithm. So, the representation of solutions in task scheduling for each chromosome consists of VMID like VM1, VM2, VM3 and ID like TS1, TS2, TS3 for each task to be executed on these VM. Each VM and the executed tasks on it are encoded into the binary bit (e.g. VM2:- TS3-TS5-TS8 →[0010—0011-0101-1000]) [8].
2) *The Fitness Function Representation:* The principle objective of task scheduling in the Cloud computing is to lessen completion time for execution all jobs on the available resources. Consequently, the completion time of undertaking task T_i on VM_j as CT_{ij} is defined using equation "(1)":

$$\text{Completion Time} = CTmax\,[i, j] \quad i \in T,\ i = 1,2,3,.....n \quad j \in VM,\ j = 1,2,3,.....m \qquad (1)$$

Where CTmax denotes maximum time for complete Task i on VM_j. n and m denote the number of jobs and virtual machines respectively.

So, to reduce the completion time which may be denoted as CTmax, the execution time of every challenge for each virtual machine have to be calculated for the scheduling purpose. If the processing velocity of virtual machine VM_j is PS_j, then the processing time for task Pi can be calculated by equation "(2)":

$$P_{ij} = \frac{Ci}{Psj} \qquad (2)$$

Where P_{ij}, the processing time for task Pi on VMj and Ci computational complexity of task Pi. The processing time of each task in the virtual machine can be calculated by equation "(3)":

$$P_j = \sum_{i=1}^{n} Pij \qquad (3)$$

3) *Selection Process:* Tournament choice is computationally extra efficient and extra amenable to parallel implementation. Consequently, the proposed TS-GA algorithm, tournament selection is used to overcome the challenge of the population size. Two individuals are chosen at random from the population. A random quantity r is then chosen between zero and one. If r < k (where k is a parameter, as an example, 0.65), the more healthy of the 2 individuals is chosen to be a parent; in any other case the much less suit person is chosen. The non-chosen individuals are then lower back to the authentic population and may be decided on once more.
4) *Crossover:* Within the proposed TS-GA algorithm, the new crossover has been used differently from the used crossover inside the original GA. Therefore; two chromosomes which might be decided on to crossover technique to generate offspring will be considered as offspring also. So, the proposed crossover produces four youngsters as shown in figure 2. After that, the two first-class kids are chosen from these.
5) *Initialize Subpopulation:* After each generation, subpopulations (i.e., new populations after crossover) are introduced into original populations (i.e., parents). This step can decorate the variety of population. This concept is added by proposed tournament selection genetic algorithm.

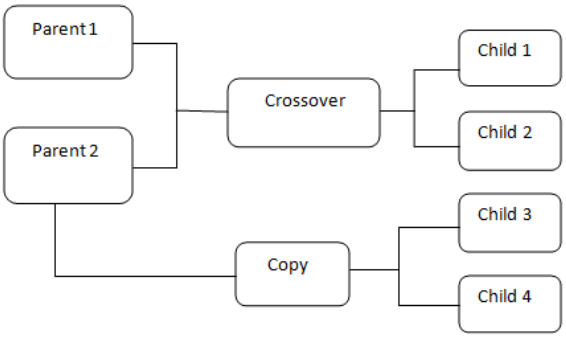

Figure 2. Crossover Process

6) Keep Optimized Solution: There may be a solution that would fulfil suitable fitness characteristic; however it isn't selected throughout the crossover method. By the use of proposed TS-GA algorithm, this solution isn't always eliminated from the population, however its miles chosen and introduced to the population while subsequent iteration is began. This step is considered as true step as a number of the iterations can generate the optimized solution. Usually, in keeping with the changed Tournament selection Genetic algorithm, a hard and fast of modifications have been brought. Those adjustments are as follows:-

a) The tournament is used as opposed to the roulette wheel in the selection system to pick the first-class solution.
b) The solutions no longer chosen within the selection method are taken into consideration and delivered to the new population. This could help in generating the exceptional solution inside the subsequent generations.
c) The new crossover is delivered by way of thinking about parents individuals as new child (see Figure 2)
d) After each new release, subpopulations (i.e., new populations after crossover) are introduced into old populations (i.e., parents).

IV. PERFORMANCE EVALUATION

In this section, the experimental assessment of the proposed Tournament selection Genetic algorithm on the default GA and Round-Robin algorithms is provided. This section begins by describing the experimental environment.

A. The Experimental Environment

CloudSim is an extensible, new, generalized simulation toolkit that enables modelling and simulation of cloud computing systems and application provisioning environments in which it is released via the Cloud Computing and distributed structures Laboratory, university of Melbourne [4]. The CloudSim toolkit supports both system and behaviour modelling of cloud system components such as data centres, virtual machines (VMs), and resource provisioning policies [3]. In line with cloudsim, the consumer attempts to publish his requests in the shape of cloudlets.

Each cloudlet has the properties of report size, the quantity of commands to be accomplished, and many others. Those cloudlets may be submitted to the broking to time table onto vms in line with scheduling [7]. Cloudsim has an advantage of the building of broker driven guidelines. The described magnificence in cloudsim, vm, represents the virtual system which can be created at the hosts. Introduction of hosts depends on the broker where it allocates each vm to the exclusive host. Datacenter has the functionality to hold a most variety of hosts and the broker can dynamically trade the setup of hosts and vms (see figure 3)

B. Experimental Results

By using cloudsim toolkit, the proposed Tournament selection Genetic algorithm is applied, and a comparative examine has been made amongst three algorithms; round-robin (RR), the default GA, and the proposed TS-GA algorithms. Three parameters are taken into consideration to evaluate the overall performance. These parameters are the completion time, execution cost, and resource utilization.

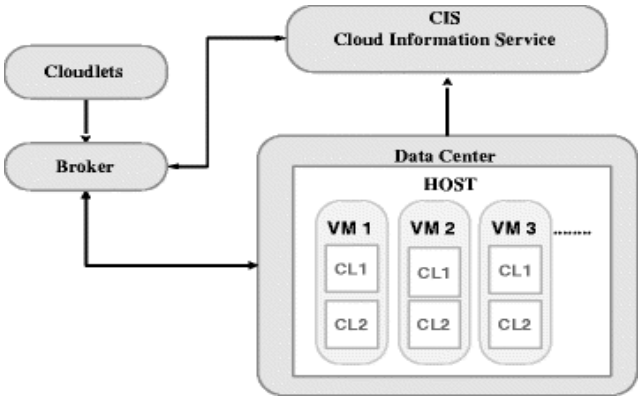

Figure 3. CloudSim Behaviour

1) Completion Time: The following table represents the completion time of three different algorithms named Round-Robin (RR), default GA, Proposed TS-GA using 8 Vms in CloudSim Toolkit. This table also provides the improvement of proposed TS-GA over RR and Default GA.

Table I. The Completion Time of RR, GA, and TS-GA Algorithms

Number of Task	RR	GA	Proposed TS-GA	Improve Proposed TS-GA vs. RR	Improve Proposed TS-GA vs. GA	No. Of VM
50	578.77	520.21	286.12	50.56 %	45.00 %	8
100	954.37	912.90	513.7	46.17 %	43.73 %	8
150	1329.97	1305.58	741.28	44.26 %	43.22 %	8
200	1705.57	1698.26	968.86	43.19 %	42.95 %	8

Following is a bar graph to show the completion times of Round-Robin (RR), default GA, Proposed TS-GA using 8 Vms.

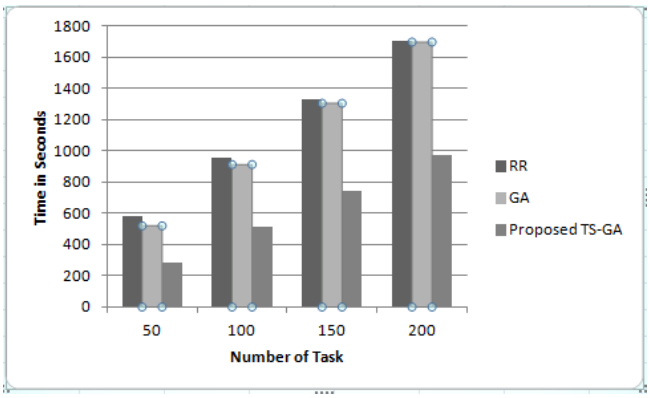

Figure 4. The comparison completion time of three algorithms RR, GA and TS-GA

According to the results shown in figure (4), its miles discovered that the completion time of the proposed TS-GA

algorithm is reduced by means of (43.725%) and (46.05%) approximately the default GA, and RR algorithms respectively.

2) *Execution Cost:* The total cost of execution of all tasks on the available virtual machines is calculated by equation "(4)":

$$Total\ cost = \frac{Task\ lengt\ h * Cost\ per\ seconds}{VM\ mips} + processing\ Cos \quad (4)$$

Table 2 represents the execution cost of Round-Robin (RR), default GA, Proposed TS-GA using 8 Vms.

Table II. The Execution Cost of RR, GA AND TS-GA Algorithms

Number of Task	RR	GA	Proposed TS-GA	Improve Proposed TS-GA vs. RR	Improve Proposed TS-GA vs. GA	No. Of VM
50	8543.22	7813.46	6603.88	22.70 %	15.48 %	8
100	14515.05	14118.87	12618.22	13.07 %	10.63 %	8
150	20486.88	20424.28	18632.56	9.05 %	8.77 %	8
200	26458.71	26729.69	24646.9	6.85 %	7.80 %	8

Figure (5) compares the execution cost of three algorithms namely Round-Robin (RR), default GA, Proposed TS-GA using 8 Vms.

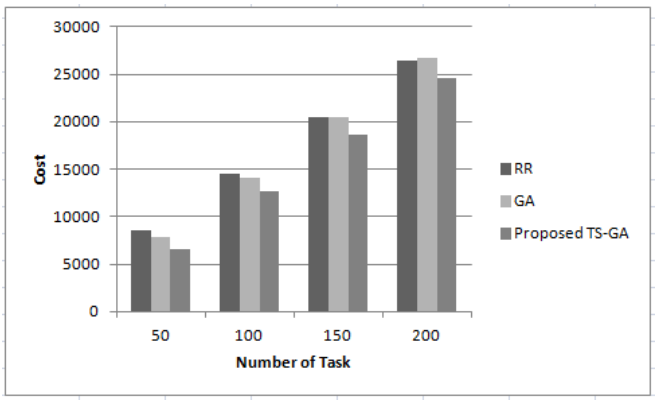

Figure 5. The comparison cost of three algorithms RR, GA and TS-GA

Consistent with the outcomes in figure (5), the cost of the Proposed TS-GA is decreased with the aid of (10.67%) and (12.92%) relative to the default GA and RR algorithms respectively.

3) *Resource Utilization:* On the opposite side, the utilization of resources represents the ratio between the entire busy time of digital gadget and the overall end execution time of the parallel utility. It can be calculated as equation "(5)":

$$Utilization = \frac{final\ VMs\ available\ time}{\#VMs * schedule\ time} * 100 \quad (5)$$

Table 3 and Figure (6) represents the resource utilization of RR, default GA and the proposed Tournament selection Genetic algorithm algorithms using 8 VMs.

Table III. The Resource Utilization of RR, GA and TS-GA Algorithms

Number of Task	RR	GA	Proposed TS-GA	Improve Proposed TS-GA vs. RR	Improve Proposed TS-GA vs. GA	No. Of VM
50	44.36	47.88	84.43	47.46 %	43.29 %	8
100	52.04	39.7	75.78	31.33 %	47.61 %	8
150	59.72	42.08	84.13	29.01 %	49.98 %	8
200	67.4	40.18	98.48	31.56 %	60.00 %	8

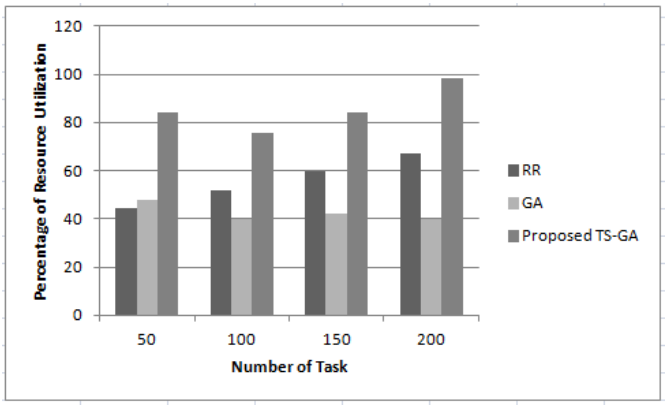

Figure 6. Resource Utilization of RR, GA AND TS-GA Algorithms

According to the effects in figure (6), its miles determined that the resource utilization of the proposed TS-GA algorithm is improved by means of (50.22%) and (34.84%) relative to the default GA and RR algorithms respectively.

V. CONCLUSION AND FUTURE WORK

This paper proposes an improved genetic algorithm named Tournament Selection Genetic Algorithm for task scheduling hassle in the cloud computing surroundings. The proposed set of rules objectives to minimize of completion time and cost, and maximize useful resource usage. The completion time of tasks for the proposed TS-GA algorithm is reduced by (43.725%) and (46.05%) about the default GA, and RR algorithms, respectively. The execution price of the proposed TS-GA algorithm is decreased with the aid of (10.67%) and (12.92%) approximately the default GA and RR algorithms respectively. The resource utilization of the proposed TS-GA algorithm is improved by way of (50.22%) and (34.84%) about the default GA and RR algorithms, respectively.

In future, the proposed algorithm may be extended to add possibility dynamic feature of vms via run GA. Furthermore, more parameters can be brought based at the customer's necessities.

VI. REFRENCES

[1] Agarwal, M., & Srivastava, D. G. (2016). A Genetic Algorithm inspired task scheduling in Cloud Computing. International Conference on Computing,Communication and Automation, (pp. 364-367).

[2] Brent, R. (1989). Efficient implementation of the first-fit strategy for dynamic storage allocation. ACM transactions on Programming Langauages and Systems(TOPLAS) , 11, 388-403.

[3] Calheiros, R. N., Ranjan, R., Beloglazov, A., Rose, C. A., & Buyya, R. (2011, August 24). CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. John Wiley & Sons, Ltd. , 23-50.

[4] Devi, R., & Sujan, S. (2014). A Survey on Application of Cloudsim Toolkit. International Journal of Innovative Research in Science, , 3 (6), 13146-13153.

[5] Ge, J., & Yuan, Y. (2013). Research of cloud computing task scheduling algorithm based on improved genetic algorithm. Applied Mechanics and Materials , 2426-2429.

[6] Ge, Y., & Wei, G. (2010). GA-based task scheduler for the cloud computing systems. Web Information Systems and Mining (WISM '10). 2, pp. 181-186. IEEE.

[7] Goyal, T., Singh, A., & Agrawal, A. (2012). Cloudsim: simulator for cloud computing infrastructure and modeling . International Conference on modeling,optimization and computing (ICMOC-2012), (pp. 3566 – 3572).

[8] Hamad, S. A., & Omara, F. A. (2016). Genetic-Based Task Scheduling Algorithm in Cloud Computing Environment. International Journal of Advanced computer Science and Applications , 7 (4), 550-556.

[9] Kalra, M., & Singh, S. (2014). Scheduling of Independent Tasks in Cloud Computing Using Modified Genetic Algorithm. Computational Intelligence and Communication Networks (CICN), (pp. 565-569).

[10] Kaur, R., & Kinger, S. (2014). Enhanced Genetic Algorithm based Task Scheduling in Cloud Computing. International Journal of Computer Applications , 101.

[11] Kaur, S., & Verma, A. (2012). An Efficient Approach to Genetic Algorithm for Task Scheduling in Cloud Computing Environment. I.J. Information Technology and Computer Science , 10, 74-79.

[12] Kumar, P., & Verma, A. (2012). Scheduling using improved genetic algorithm in cloud computing for independent tasks. International conference on Advances in Computing, Communications and Informatics, (pp. 137-142). Chennai.

[13] Madureira, A., Ramos, C., & Carmo Silva, S. d. (n.d.). Using Genetic Algorithms for Dynamic Scheduling.

[14] Raj, J., & Thomas, R. (2013). Genetic based scheduling in grid systems: A survey. International Conference on Computer Communication and Informatics (ICCCI) , (pp. 1-4).

[15] Ravichandran, S., & Naganathan, D. (2013). Dynamic Scheduling of Data Using Genetic Algorithm in Cloud Computing. International Journal of Computing Algorithm , 2, 127-133.

[16] Sindhu, S., & Mukherjee, S. (2011). Efficient Task Scheduling Algorithms for Cloud Computing Environment. International Conference on High Performance Architecture and Grid Computing (HPAGC-2011), 169, pp. 79-83.

[17] Thyagrajan, K., Vasu, S., & Harsha, S. (2013). A Model for an Optimal Approach for Job Scheduling in Cloud Computing . International Journal of Enginnering Research and Technology .

[18] Vignesh, V., & Kumar, K. S. (2013). Resource management and scheduling in cloud environment. International Journal of Sceintific and Research Publications , 3.

[19] Zheng, Z., Wang, R., Zhong, H., & Zhang, X. (2011). An approach for cloud resource scheduling based on parallel Genetic Algorithm. International conference on Computer Research and Development (ICCRD), (pp. 444-447).