Novel Approach of Scientific Workflow scheduling By hybrid Swarm intelligence for resource Utilization

Harpreet Kaur, Dr. Manish Mahajan

Department of Computer Science and Engineering, CGC COE Landran, Mohali, Punjab

Abstract - Workflow scheduling in scientific computing systems is one of the most challenging problems that focuses on satisfying user-defined quality of service requirements while minimizing the workflow execution cost. So, to reduce the cost we use cloud environment. In cloud environment, resources will increase but it's utilization is another challenge while using cloud environment. In this paper, we use intelligence optimization ant colony optimization which is initialized by Pareto distribution. ACO is used to converge the decision of Virtual Machine (VM) migration by its convergence to minima of cost and time. In our experiments, we use Total execution time(TET) and Total execution cost (TEC) which ACO_PSO shows significance perform when compare to existing Genetic algorithm and PSO.

Keywords - Workflow Scheduling, Ant colony optimization, Cloud computing

I. INTRODUCTION

The distributed data sources use is involved in domains such as neuroscience, structural biology, and high-energy physics in the modern scientific collaborative experiments. As a result, their datasets analysis is represented and structuring it as workflows scientifically [1]. Usually, these scientific workflows require the huge data amount process and intensive computational activities. A workflow scientific management system [14] is utilized to manage these experiments scientifically by the integration and orchestration details hiding inherent while cloud service providers provides workflows execution on distributed resources. A distributed computing new paradigm is cloud computing which delivers platform, software (application), and infrastructure as services. As services based on subscription these services are made available in a pay-as-you-go model for consumers [3, 2]. Dynamically user applications provisioning helped by cloud computing at specified locations as many computing resources as per the requirement. The storage locations are chosen by applications for hosting their data at global locations. For the efficient and cost effective scheduling of the tasks and applications data onto these environments of cloud computing, different policies are there in the application schedulers who vary in accordance to their objective function: total minimized cost for execution, load balance on resources used while the application deadline constraints is meant, and so forth. This paper focuses on the total execution cost minimization of applications on these resources which are provided by service providers of cloud, such as GoGrid3 and Amazon. This is achievedby the utilization of a hybrid approach having Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO). A global search self-adaptive based optimization technique is Particle Swarm Optimization (PSO) which was introduced by Kennedy and Eberhart [2]. There is a similarity among algorithm proposed and other algorithms based on population like Genetic algorithms but, no direct individuals re-combination of the population and it relies on the particles social behavior. Each particle trajectory is adjusted in every generation which is based on its best position (local best) and its best particle's position (global best) in the entire population. In this concept, the stochastic particle nature increases and quickly converges for global minima with a good solution reasonably. A random search algorithm referred to as ACO algorithm, such as other evolutionary algorithms[4]. The real ant colonies behavior is imitated by ACO in nature for food search and for connecting each other by trails of pheromone laid on the travelled paths. ACO is utilized by many researchers for solving NP-hard problems like problem of traveling salesman [4], problem of coloring graph [11], problem of vehicle routing[12], and so on. This research work principle idea is that for integrating some basic cloud computing philosophies such as the computing resources heterogeneity and elasticity and for meeting the Quality of Service (QoS) user's requirements on Infrastructure as a Service (IaaS), for scientific workflows ,the resource scheduling and provisioning strategy required being there. For execution time and cost minimization, the proposed PSO algorithm but in several cases it fails in producing schedules with inferior make span and and greater costs is grater as the deadline is met [14]. Hybrid PSO (ACO-PSO) was utilized in Mobile ad-hoc network and gives better presentation in terms of power, delay, consumption, and communication cost for comparing for standard PSO. For better results if ACO-PSO optimization algorithm in an IaaS Cloud, an efficient execution plans for designing in accordance to reduced make span and pay a lower price.

II. LITERATURE REVIEW

The task scheduling having low complexity and is performance effective for heterogeneous computing has been proposed in [3] consist of two approaches to schedule a workflow task in heterogeneous environment of computing namely Process or critical path and HEFT. There work is on the similar line with slighter variations. HEFT utilizes the upward task prioritization rank; the other utilizes the upward and downward and upward ranks combination. The critical path is utilized by the latter and the tasks are assigned on the critical path for the processor which gives less EFT. Better is the HEFT in comparison with other algorithms in similar domain because of its high make span and robustness in nature. A Hybrid Heuristics- Genetic Algorithm for In Heterogeneous Processor Networks H2GS for Task Scheduling [1], a hybrid heuristic genetic algorithm in [2] is defined in which large execution costs and information retrieval incurred by user program while scheduling them taken into account the ‘execution time’ only. Further for execution time optimization, the cost springing up from statistics resources transfers in addition for execution expenses required to be taken under consideration. A particle swarm optimization (PSO) offered by creator which is based heuristic to cloud sources for time table applications which takes into account every cost computation and transmission price statistically and with current ‘Best Resource Selection’ (BRS) rules set evaluation.

An Ant Colony Optimization (ACO) algorithm is proposed in [3] for scheduling massive-scale workflows along with numerous parameters QoS. Customers get permits from this algorithm for specifying their QoS options and minimal QoS thresholds outlined to sure software. This rules set goal was for finding a solution which meets every QoS constraints and the user-preferred parameter of QoS is optimized which is based totally at the workflow scheduling traits, new seven heuristics designed by authors for the ACO technique and an adaptive scheme is proposed which lets in synthetic ants for selecting pheromones values that are heuristics based.

Architecture is advised in [4] for the automatic execution of the application based on primarily huge-scale work flow on elastic and dynamic provisioned resources of computing. Authors centered on middle rules set namely Partitioned Balanced Time Scheduling (PBTS) that estimates the minimal computing hosts variety required for executing a workflow inside end time of distinctive user. The designing of PBTS rules set for in-shaping every useful elastic resource which provisions models together with AmazonEC2 and parallel malleable utility models constitute of Map lessen. The aid potential estimated by the PBTS close to the certainly low theoretically is verified by the author.

The LDCP list is utilized in [5] based on heuristics for generating the population initially. Longest critical Dynamic path is a tri-phase heuristic list based (selection of task, selection of processor and update of 12 status). In first phase, calculating the task priority by utilizing the up ward rank. In second phase, the resultant of the processor is the selection of minimum EFT. Status updates phase and also updates the system status. GA (genetic algorithm) and LDCP are combined by H2GS. LDCP generates the schedule of higher quality as a seed utilized by it for the initial population that the customized genetic algorithm exploits. The LDCP generated schedule is nearer to an optimal schedule and as this type of schedule to the genetic algorithm is given as an input, faster will be the convergence of the algorithm. 2-D chromosomes are utilized to represent and customized operators are utilized for searching the problem space. In normalized schedule length and speedup, significant improvement is shown over Dynamic Level Scheduling and HEFT in distributed heterogeneous system. Genetic Algorithms an efficient approach for Task Scheduling in Environment of cloud computing.

A new Modified Genetic Algorithm is proposed in [6] to schedule tasks in private cloud to minimize the cost and make span. In MGA, Longest cloudlet to Fastest Processor(LCFP), 8 random schedules, and Smallest cloudlet to Fastest Processor (SCFP) are utilized for generating initial population. Two point simple swap and crossover are utilized. The good performance is given by this under the heavy loads. A Performance Effective Genetic Algorithm for Task Scheduling (PEGA) in Heterogeneous Systems proposed in [6] is an effective genetic algorithm that has the optimal results providing capability in larger space having lesser time complexity. Two parts are utilized in the representation of the direct chromosome. The b-level (upward rank) make the right half that gives the better results in context of schedule length in comparison with the population generated randomly. For quality enhancement and solution’s convergence speed, two fold cross over is utilized that executes two point and single crossover one after the other. The better schedule is provided by PEGA as concluded by the author with low time complexity and smaller make span[13]. A Hybrid heuristic-Genetic Algorithm in Heterogeneous Multi-Core System (HSCGS)for Task Scheduling in [18] presented an approach that uses a successor combination which is concerned list based on genetic algorithm and heuristic. First phase is the GA seeding method that is a method for generating the population initially by the given schedule by using the SCLS(Successor concerned list heuristic). The task priority list in SCLS was formed utilizing the upward rank. The above phase is utilized for generating the good quality schedule in the second phase which is fed into the genetic algorithm. Better results are obtained by the HSCGS as proved by the author in comparison with HEFT and DLS(Dynamic Level Scheduling) Deadline constrained Workflow Scheduling Algorithms for IaaS Service Clouds [8].

III. ALGORITHM USED

A. Particle Swarm Optimization (PSO)

Particle swarm optimization (PSO) is a modeled algorithm on intelligence of swarm which finds a solution for a problem of optimization in a search space, or the social behavior is predicted and modeled in the objectives presence. The PSO is a stochastic, computer algorithm based on population which is modeled on intelligence of swarm. Social-psychological principles based are the swarm intelligence which provides insights to the behavior socially along with its engineering applications contribution.

B. Ant Colony Optimization (ACO)

The major idea of Ant colony optimization is the problem of modeling in a graph, minimum cost while searching path from source to destination. This graph is as there are artificial walking ants on it and to find a path leading up to lesser expense. The capability of single ant having simpler behavior to find paths which are relatively less costly. In the ant colony, the result amalgamated in the global ant cooperation gives path that is less costly. The real ant inspired is the artificial ant's behavior, similarly as the pheromones trails that the ants laid on the graph edges and selecting their path based on the pheromone trails probability. Some features are additional in artificial ants as compared to real ants.

IV. SYSTEM MODEL DESCRIPTION

***Define the workflow*:** Different workflows from different scientific areas are choosen and simulated on cloud environment.

***Apply Particle Swarm Optimization Algorithm*:** PSO is a problematic optimization technique in which the most basic concept is that of particle. A particle represents an individual (i.e. fish or bird) that has the ability to move through the defined problem space and represents a candidate solution to the optimization problem.

Apply Ant Colony Optimization Algorithm: Output PSO will be given as input to ACO for finding the best path. The algorithm Ant colony optimization (ACO) is a heuristic algorithm which is based on the behavior of the ants seeking the shortest path between anthill and the location of food source. With the mechanism of positive feedback and distributed cooperation, it is proved to be a useful heuristic algorithm for solving NP hard problems.

Schedule Generation: Construct a schedule by converting a particle's position.

Evaluate Make span: Make span is evaluated to meet the given deadline.

***Evaluate Cost and time*:** For each Workflow cost in terms of money is calculated and time will be seconds.

Compare the result: Simulate and compare the result of proposed algorithm with PSO and GA

Proposed system model

Step 1: Define the workflow

Step 2: Initialize Particle (p_1,p_2,…..p_n) and starting weight a_1 and a_2.

Update velocity

$$V_x^{t+1} = V_x^t + a_1 c_1^t (pbest_x^t - U_x^t) + a_2 c_2^t (gbest - U_x^t)$$

If $(A_p < \xi)$

{

Analyze the parameters

}

else

update task by pheromone value

while $(A_b < A_p)$

$$p_{xy}^n = \left\{ \frac{J_{xy}(t)^\alpha (\eta_{xy})^\beta}{\sum (J_{xy}(t))^\alpha (\eta_{xy})^\beta} \right\}$$

Optimize α, β

end

Step 3: Initialize the number of Ants in accordance to the obtained PSO Output.

While (true)

N++ // Simulation Steps

for each resource

if (source = destination)

return

else

change in pheromone

for each path by eq. given below

$t_{xy} = (1- \rho)\, t_{xy} + \sum_n \Delta\, t_{xy}$

t_{xy}← pheromone updation on the edge from node x to node y

n← iteration

ρ- density

$\Delta\, t_{xy}$– amount of pheromone deposited

for each neighbouring path

Calculate probability by eq. given below

$p_{xy} = (t_{xy})^\alpha * (\eta_{xy})^\beta / \sum (t_{xy})^\alpha * (\eta_{xy})^\beta$

η_{xy}-is the distance inverse among the two nodes

α -parameter for controlling the of t_{xy} influence

β- parameter for controlling the η_{xy} influence

end

if (scheduled)

break;

else continue

end

end

Step 4: Evaluating and comparing the obtained parameters.

V. EXPERIMENTAL RESULTS

Workflow	TEC	TET (sec)	F-GEN	STORAGE(mb)
2	103.22	26.7	36.224	10.21
4	106.63	37.5	36.321	386.7
6	231.8	57.7	38.237	498.9
8	262.74	69.1	38.639	886.6
10	367.26	76.8	39.042	1191.1

Table 1: Performance comparison on the basis of the above given parameters using PSO approach.

Workflow	TEC	TET(sec)	F-GEN	STORAGE(mb)
2	60.51	21.8	26.268	163.4
4	109.93	37.4	32.768	344.6
6	202.99	46.3	35.066	633.2
8	219.88	67.5	37.608	881.5
10	261.53	75.6	39.048	11902.5

Table 2: Performance comparison on the basis of the above given parameters using GA approach

Workflow	TEC	TET(sec)	F-GEN	reduce the time of computation because single objective optimization which use in research work in serial way and increase the computation time.STORAGE
2	57.63	21.7	8.424	36.56
4	75.89	37.4	10.288	414.2
6	123.06	45.9	9.441	431.2
8	147.08	59.5	9.497	795.1
10	258.92	70.2	10.065	1190.1

Table 3: Performance comparison on the basis of the above given parameters using Hybrid ACO-PSO approach

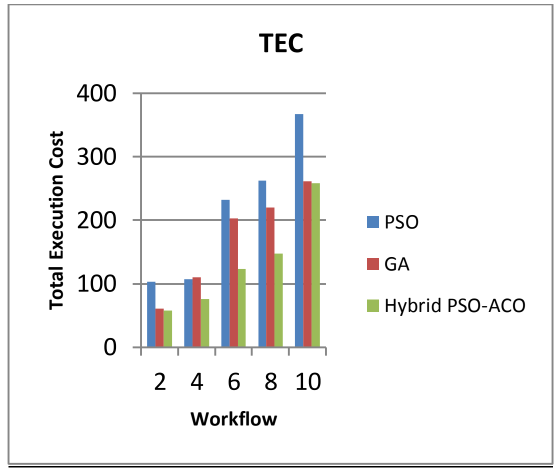

Fig.1: Performance comparison of Hybrid PSO-ACO approach with PSO and GA based on TEC

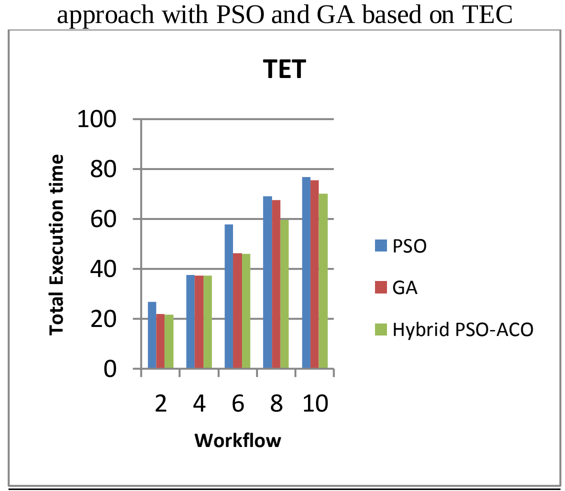

Fig.2: Performance comparison of Hybrid PSO-ACO approach with PSO and GA based on TET

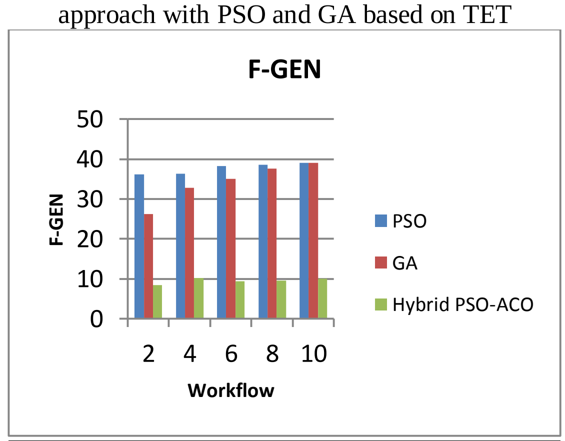

Fig.3: Performance comparison of Hybrid PSO-ACO approach with PSO and GA based on F-GEN

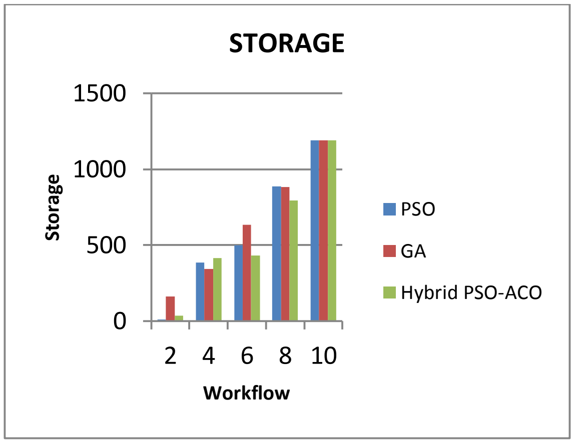

Fig.4: Performance comparison of Hybrid PSO-ACO approach with PSO and GA based on Storage

VI. CONCLUSION

In our experiments, we work on different workflows genome, cyber etc on TET and TEC parameter in different virtual machine or ensemble size. In this paper, we use two to twenty ensembles size and optimize by genetic algorithm and ant colony optimization. In experiment results, ACO_PSO reduces the average TET and TEC in different workflow. So, we concluded that ACO_PSO optimize and coverage workflow scheduling in cloud scenario. In future this work enhance on hybrid optimization and multi objective optimization because both are

VII. REFERENCES

[1]. M. I. Daoud and N. N. Kharma, "A hybrid heuristic-genetic algorithm for task scheduling in heterogeneous processor networks", Journal of Parallel and Distributed Computing, Elsevier, vol. 71, no. 11, pp. 1518–1531, 2011 .

[2]. Y. Xu, K. Li, J. Hu and K. Li, "A genetic algorithm for task scheduling on heterogeneous computing systems using multiple priority queues", Information Sciences, Elsevier, vol. 270, pp. 255 – 287, 2014.

[3]. H. Topcuoglu, S. Hariri and M.Y. Wu, "Performance-effective and low-complexity task scheduling for heterogeneous computing", IEEE Transactions on Parallel and Distributed Systems, vol. 13, no. 3, pp. 260-74, March, 2002. 24

[4]. S. Smanchat and K. Viriyapant, "Taxonomies of workflow scheduling problem and techniques in the cloud", Future Generation Computer Systems, Elsevier, vol. 52, pp.1-12, November, 2015.

[5]. S. Kaur and A. Verma, "An efficient approach to genetic algorithm for task scheduling in cloud computing environment", International Journal of Information Technology and Computer Science (IJITCS), vol. 4, no. 10, pp. 74, September, 2012.

[6]. S. G. Ahmad, E. U. Munir, and M. W. Nisar, "PEGA: A performance effective genetic algorithm for task scheduling in heterogeneous systems", IEEE 14th International Conference on High Performance Computing, pp. 1082–1087, 2012.

[7]. C. Wang, J. Gu, Y. Wang, and T. Zhao, "A hybrid heuristic-genetic algorithm for task scheduling in heterogeneous multi-core system", in Algorithms and Architectures for Parallel Processing (Y. Xiang, I. Stojmenovic, B. Apduhan, G. Wang, K. Nakano, and A. Zomaya, eds.), Springer Berlin Heidelberg ,vol. 7439 of Lecture Notes in Computer Science, pp. 153–170, 2012.

[8]. S. Abrishami, M. Naghibzadeh,and D.H. Epema, "Deadline-constrained workflow scheduling algorithms for Infrastructure as a Service Clouds", Future Generation Computer Systems, Elsevier, vol. 29, no. 1, pp. 158-69, January, 2013. 25